Comments and Observations on "The Science and Superstition of Quantitative Risk Assessment"

by Arthur Barondes

(Editor's Note: The following is a critique of the article "The Science and Superstition of Quantitative Risk Assessment" by Andrew Rae, John McDermid and Rob Alexander, which ran in the July-August 2012 issue of Journal of System Safety)

QRA Superstition. Not fit for purpose. Really? The article by our Yorkshire colleagues goes back and forth in describing various pros and mostly cons of Quantitative Risk Assessments (QRAs) before settling on: safety practitioners' time "may be better spent on activities other than QRA," such as "thinking about how to make a system safer." One is left to ponder whether "thinking about safety" is more "fit for purpose" than QRAs. More important, one wonders why "thinking" and QRAs are offered as either-or alternatives when QRAs entail extensive thought processes and are not the sole basis for risk-informed safety decisions. Based on our extensive experience with QRAs for engineered systems, we challenge many of the authors assertions and research, and make some counter observations on QRA limitations.

Clarification

Before examining the authors' arguments, some clarification is needed on how the authors' structured their case against QRA fit for purpose (defined as: "good enough to do the job it was designed to do"). They describe three forms of QRA applications that vary in scope, and then proceed to attack the use of QRAs on the basis of their broadest (and most unlikely) application, viz., assessing all system safety risks, "the top number" — the aggregated measure of total system risk" as in "determining total risk [for]…all [emphasis added] of the real world consequences." They generalize their "fit for purpose" arguments on the basis of that broadest application — one that, in our view, is largely impracticable and unreasonable for engineered systems. Indeed, we would agree that as a practical matter, QRAs are not fit for — nor, as far as we know, used for — such far-reaching assessments. But we question the logic of generalizing from that unrealistic application to an all-encompassing denigration of QRA "fit for purpose." In particular, we find QRAs to be applicable and fit for purpose for assessing risks of specific high consequence, rare events that warrant their high costs in time and money, as in the Nuclear Regulatory Commission (NRC) WASH-1400 that took four years and cost $4M (1975 dollars).\(^1\)

Admittedly, late in the paper, the authors give that type of QRA a qualified "free pass." They concede that such focused QRAs may be "sufficiently accurate and precise...for specific outcomes arising through specific mechanisms, as many of the challenges to QRA [generated by the authors] do not apply in these cases [emphasis added]." But the damage is done. We are left to conclude — perhaps erroneously — that the authors are presenting a "stalking horse" to stimulate debate on the "fit for purpose" of QRAs. Unfortunately, in the process of generalizing on impracticable applications, they have cast unwarranted aspersions on what others see as "tremendous progress toward rational decision making."\(^2\)

Observations

Recognizing that the authors may be seeking a solution to a non-problem, we make a number of selected observations to set the record straight. To begin, we applaud their skepticism and some of the issues they raise. However, we find it important to note that their attacks on QRA "fit for purpose" are based, for the most part, on examples of shortcomings not in the QRA methodology, but in the execution. They present an overview of the QRA methodology from the "30-year-old 1983 "Red Book\(^2\) [tailored to adverse health effects, e.g., cancer, rather than engineered systems] with the authors' closing critical observation: "very little information exists on the validity
of the techniques…and good reason to view the results…with skepticism.*

We do not find that dearth of information. As an example, the 34-year old (1978) NRC-chartered Lewis Review Group (LRG) of WASH-1400 (cited by the authors) — an in-depth peer review — includes an entire page (p. 45) of examples of techniques and their use in the regulatory process. A few are quoted here:

Use of the techniques was anticipated in several applications made before the Study was completed. A probabilistic study of aircraft crashes has led to a criterion limiting acceptability of sites for nuclear power plants to areas far enough from airports. The earlier reviews (1972-73) of anticipated transients without scram (ATWS) started from a requirement that these events have a probability of less than $10^{-6}$ and assessed the ability of reactor designs to meet this. Other uses of probabilistic methods have been made with varying success in increasing numbers since issuance of the Reactor Safety Study. A few examples are: more refined methods of seismic analysis based on probability of occurrence of earthquakes and probability that structures can withstand assumed earthquake stresses, performance of backup power systems, the effects of overpressure and other transients on reactor pressure vessels, the probability of piping failure, and the effects of missiles generated by breakup of a turbine.

That aside, we argue against the authors' defined "primary claim" for QRAs, viz., that they must have sufficient accuracy and precision for "aggregated…total system risk." As described earlier, each QRA should have sufficient accuracy for its specific purpose. That is rarely, if ever, assessing aggregated risk. Rather, it is assessing the risk of a single specific end state (mishap), e.g., core meltdown, and often ordering risk contributors to support risk management. The use of a single end state is exemplified with the authors' use of WASH-1400. That pioneering 1975 QRA examined two reactors to assess a specific end state: core meltdown and consequent health effects — not "aggregated…total system risk."

Looking more closely at WASH-1400 and the LRG report, we form a different picture than the authors do. Whereas the LRG severely criticized parts of that QRA's execution, it endorsed the methodology and advocated its wider use. The LRG report states (p. vii):

We do find that the methodology, which was an important advance over earlier methodologies applied to reactor risks, is sound, and should be developed and used more widely under circumstances in which there is an adequate data base or sufficient technical expertise to insert credible subjective probabilities into the calculations.

As for the values of the risk estimators and their uncertainty bands, the LRG said, "We are unable to determine whether the absolute probabilities of accident sequences in WASH-1400 are high or low, but we believe that the [stated] error bounds on those estimates are, in general, greatly understated," — i.e., optimistic. However, with the benefit of hindsight, we note that subsequent years of real-world U.S. reactor experience that includes the Three Mile Island accident — "extant evidence" — shows core meltdown frequency within the WASH-1400 upper uncertainty bound. (Note: The authors erroneously claim that "uncertainty ranges were not provided" in WASH-1400, and that the LRG "raised concern…that the uncertainty…was unstated...[and] unknown." They also refer to claims that the probability estimates were pessimistic.)

The authors also cast doubt on the QRA methodology by citing that "one of the [seven LRG] authors...questioned whether the methods could ever provide estimates with sufficient confidence." But that is not an accurate interpretation of Dr. Frank von Hippel's reservation. As stated in the LRG report, he "...questions whether, for a system as complex as a nuclear power plant, the methodology can be implemented to give such a high level of confidence, that the summed probability of many known and unknown accident sequences leading to an end point such as a core melt is well below the limit set by experience [emphasis added]." Note that the reservation is for accepting probabilities that are not consistent with experience — not for the QRA methodology, per se. The NRC translated Dr. von Hippel's concern into "quantitative risk assessment techniques...should not be used to estimate absolute values of probabilities of failure of subsystems unless an adequate data base exists, and it is possible either to quantify the uncertainties or to support a conservative analysis." That was in the late 1970s, in the typewriter days, before the widespread use of computers and, in particular, bench-marked fast-running simulation codes to fill data voids.

Next, we examine the authors' arguments on causal models. Whereas comprehensive hazard identification is part of the QRA methodology and essential for executing the methodology, we would argue that hazards are most apt to be identified using a methodology that actively seeks them in a structured program. The QRA methodology includes overarching qualitative fault trees that have the specific purpose of identifying all hazards and initiating events that might lead to the undesired end state (top event). These Rasmussen-inspired Master Logic Diagrams (MLDs) are peer reviewed and updated for newly discovered hazards. This is a "process of systematically
analyzing the sources of risk" that the authors claim is an alternative to QRA. While there is no guarantee that every hazard will have been identified by any process, the MLD approach requires a best effort to assure completeness. (Note: There is no reason to believe that QRAs must eschew qualitative or deterministic methods.)

We also take issue with the question-and-answer sequence on the utility of a QRA that produces an inaccurate "top number" (a loaded question to begin with). In something of a non sequitur, the authors cite a paper by Apostolakis6 (now a NRC Commissioner), saying he "chose specifically not to defend the quantification of aggregate risk." In our reading of his paper, we do not find any substantiation for that observation. Indeed, the paper does not contain the phrase "aggregate risk" — perhaps because QRAs are not suited to that purpose. However, the paper does contain numerous references to the productive use of Quantitative Risk Assessments for high-consequence, rare events.

Continuing with the Apostolakis paper, the authors mistakenly state that he "concentrates on identifying opportunities for design improvements or further assurance." In fact, he concentrates on safety programs that exhibit "defense-in-depth," as in, "It is my thesis that QRAs should be viewed as an additional tool in safety analysis that improves safety-related decision making... [although the tool is not perfect] it represents tremendous progress toward rational decision making." And, "I wish to make one thing very clear: QRA results are never the sole basis for decision making by responsible groups... safety-related decision making is risk-informed, not risk-based. The requirements of the traditional safety analysis...are largely intact." Since we might be viewed as biased in our reading, we strongly recommend that those interested in the subject read the Apostolakis paper for themselves (see link in footnote 2).

Next, the authors identify a number of potential problem areas that, although real and worthy of serious consideration, are extraneous to the issue of QRA methodology "fit for purpose." They deal with analyst and reviewer bias, competence, integrity and the like, as well as management considerations, such as effective communication, separation of risk assessment and risk management. The "reviewer" issues are interesting, not because they are germane to evaluating QRA methodology, but because they introduce the reader to the otherwise neglected quality assurance role of structured peer involvement and review by independent experts embedded in the methodology. Whereas peer review poses its own problems in execution, it has long been accepted in the scientific community as "the best method of assuring the technical credibility of such a complex undertaking."7

QRA Limitations

We could go on with more comments and observations, but we find these sufficient to question the QRA "fit for purpose" paper — or, in the authors' words, "good reason to view the results...with skepticism." Instead, accepting that QRAs are not perfect, we offer what we see as some real programmatic limitations — as in cost, schedule and performance — in the execution of what we see as a fit for purpose QRA methodology.

First, QRA execution is time-consuming and expensive. Again, WASH-1400 was a four-year, $4M effort in the 1970s. This limits the application of QRAs to risks that are commensurate with the QRA costs. This usually means risks of high-consequence, rare events.

Second, QRA execution must get the physics and statistics right. This is easier said than done. This can require material properties outside the range of empirical data and extreme environments not normally encountered. The statistics must produce risk estimators based on a host of random variables expressed as distributions with confidence intervals and correlations in events. The statistics must also accommodate uncertainties as they apply to the risk estimators.

Third, QRA execution can face critical data voids and uncertainties. These can be "show stoppers" or expensive "show extenders." Accommodating them can involve test programs and structured expert elicitations that add cost and can affect schedule. Worse, aleatory and epistemic uncertainties can be too great to order risk contributors with statistical significance.

Fourth, required peer involvement during the course of QRA requires disinterested peers. These can be hard to find. Further, their reviews can lead to extensive and expensive re-work.

Fifth, QRAs are not intended to answer the question, "How safe is safe enough?" They provide quantitative risks estimators (usually central values from risk distributions) and their uncertainty bands. "How safe is safe enough?" is answered by management in terms of willingness to accept risk. QRAs can also identify risk mitigation actions for consideration by management.

About the Author

Arthur Barondes is a principal at Analytics International Corp., based in Alexandria, Virginia.
6 Apostolakis, G. E., op. cit.
7 Lewis, op. cit., p. vii.