UCRL- 92125
PREPRINT
CONF-850809--26

UCRL--92125

DE85 008295

Use of Reliability Data for QA Program Evaluation

Sergio B. Guarro

This paper was prepared for submittal to
8th International Conference on
Structural Mechanics in Reactor Technology
Brussels, Belgium

August 19-23, 1985

Lawrence Livermore National Laboratory

Abstract

Possible analytical approaches for evaluation of the effectiveness of QA programs implemented in the operation of U.S. commercial nuclear power plants are discussed. These approaches may be based on key plant component performance comparisons, correlation models, or comprehensive cost-benefit evaluation frameworks. As plant availability and reliability data must be used to quantify the models, the quality of these data conditions the amount of information that can ultimately be extracted. The potential impact of uncertainties in the data must be considered carefully, especially before application of the more complex models.

1. Introduction

The assurance of quality in power plant construction and operation has always been an important issue in the nuclear industry and regulatory arena. Studies have recently been performed to qualitatively evaluate the implementation of QA requirements by utilities during the plant commissioning and construction stages [1]. The question has also been raised of whether it may be possible to develop a valid framework for quantitative evaluation of quality assurance and quality control practices affecting construction and operation [2]. Such a quantification, if at all possible, would allow decision makers to review, on a cost-benefit evaluation basis, current regulatory requirements imposed on U.S. nuclear utilities for development and application of comprehensive QA programs.

These topics are addressed here, with exclusive focus on the QA practices that apply to the plant operation stages. In this context any proposed approach will have to make use of the currently available statistical data pertaining to plant reliability or safety performance. This may be a critically limiting factor at the present time, as the data are far from encompassing complete sets of information. In the following we discuss some of the insights that were gained by trial use of different evaluation approaches and available databases.

2. Possible Analytical Approaches

Evaluation of the effectiveness of a complex combination of operational activities as those included under the generic definition of "nuclear QA Program" is a very complex task. While Appendix B to 10 CFR 50 defines these activities and the associated requirements in a broad sense, actual interpretation of the requirements varies considerably from one utility to another. This often makes the level of quality control and quality assurance applied to the same type of component in different plants to be non uniform, and hinders the development of an evaluation framework that were to be based on generic — i.e. non-plant specific — information. Non trivial problems also exist in the definition of the information that one may want to obtain from any such a framework. A nuclear quality assurance program can in fact be assumed to have an impact on many different areas associated with commercial plant operation. Table I gives a schematic picture of some of the most important areas that are impacted. The central column of the Table indicates direct cost areas that can be associated with a QA program, whereas the last column indicates areas where some sort of induced benefit can be expected to be produced.

Development of a global evaluation framework would require computation of all the significant costs and benefits that may be directly or indirectly associated with the identifiable program impact areas and their reduction to some measurable dollar-equivalent baseline units. This may not be presently feasible. Factors like public confidence or, traceability of performance, maintenance and repair records cannot be easily defined and measured in quantitative terms. For quantitative evaluation, one would actually have to "measure" changes in these factors that could be associated with different hypothesized QA levels of implementation.

An evaluation based on development of a model covering the first two impact areas listed in Table I is in theory possible, but would require a great amount of effort. It can also be expected that the indications obtained thereof would be subject to very large uncertainties. In principle such a model could be constructed similarly to a PRA (Probabilistic Risk Assessment) model, and then used to calculate variations in plant risk and availability indexes due to real or hypothesized changes in implemented QA practices. To allow a computation of this kind, one would have first to determine what variations would be produced by such changes at the basic component failure rate level. To obtain a complete quantification, the basic failure rate variations would have then to be input and propagated through the model for all important components. Availability of specialized and comprehensive component failure data becomes thus, as one may have expected, a crucial and limiting factor. Unfortunately, the currently available data sources, some of which are briefly reviewed in Section 3, simply do not contain the quantity and quality of information that is needed for a meaningful implementation of the model.

If development of a comprehensive model appears presently problematic, evaluations and analyses performed within a more limited context can be realistically pursued. Component performance variations due to different QA requirements may be quantifiable in selected cases and under simplifying assumptions. A comparison of the performance of safety-related plant components with non-safety-related components of similar characteristics, as a possible indicator of the effectiveness of the current nuclear QA program, has been suggested in the past by the ACRS (Advisory Committee on Reactor Safeguards) [3]. Section 4 discusses results collected by the author and others in performing limited analyses of this kind. The problem with this approach is that it affords comparisons at the component level, but it does not tell much on the effects that QA may have at the whole plant level.

To gain some insight at this higher level one may wish to look for ways of relating available QA practice indicators with plant performance indicators. The latter are readily found in the form of capacity factors and other similar parameters, as discussed in some more detail in Section 3. On the other hand, detailed records of violations by plant operators of QA program requirements set forth in 10 CFR 50 Appendix B or in other licensing documents are maintained by the U.S. NRC Office of Inspection and Enforcement (IE) . These records are collected in the so called "766 File", which is also briefly discussed in Section 3. Using these sets of data, it is possible to perform an empirical analysis and determine if, in any group of operating plants, a correlation can be observed between plant performance indicators — such as number of shutdowns, or overall capacity factors — and the recorded number of QA program violations — which is an indicator of the QA level of implementation in each of the given plants. Results from an analysis like this are presented in Section 5.

3. Available Data Bases

The analytical techniques discussed in the previous section require all use of statistical data. Looking more closely we can see that the usable information can be classified in one of the following three categories:

A) Component reliability data

B) Plant availability data

C) Plant regulatory performance records

Category A comprises compilations of historical records of equipment performance. These data allow derivation of component failure rates and failure on demand probabilities. Examples of this type of data base are the LER (Licensee Event Report) File, the NPRDS (Nuclear Plant Reliability Data System) and the IPRDS (In-Plant Reliablity Data System). From the point of view of comparing performance of safety - vs. non-safety-related components, all of the above data bases present problems. The LER File, established and maintained by the U.S. NRC, was not designed as a source of reliability data: it contains no information on component populations, exposure times and demand rates. In addition, failures of non safety related equipment are not reportable under this system. The NPRDS, currently maintained by the Institute for Nuclear Power Operations (INPO), was inititated by voluntary utility participation and, although currently being improved through INPO's efforts, still suffers from under-reporting and incompleteness in some of the areas that were also indicated as problem areas in the LER File system. More detailed data are currently being made available through IPRDS, a system activated fairly recently through the sponsorship of the U.S. NRC and the American National Standards Institute (ANSI). This system is designed to compile data directly from plant component maintainance and repair files. Component population data are available in this system, but the information on exposure times and demand rate is presently still based on rather gross estimates, which results in large unquantified uncertainty in the calculated failure rates. The newness of the system, which presently covers only a few years and a restricted number of plants, also introduces statistical uncertainties in the estimations, with 90% confidence intervals in failure rate values often spanning two order of magnitudes. We finally note that specialized component reliability data banks are compiled and maintained by nuclear system vendors and manufacturers. These data, however, are usually regarded as proprietary information, and therefore as not generally accessible for public use.

Plant availability data — Category B — are available in different forms through several sources, including the LER and NPRDS files. The best public source for these data, however, is in the monthly published issues of NUREG-0020 [4], which are compiled on the basis of operating reports submitted by all licensees to the NRC office of Management and Program Analysis. The information made available for every given month includes the number of hours of reactor criticality, electrical generation and forced outages, various availability and capability factors, dates of scheduled shutdowns, and causes of all outages.

Plant regulatory performance records are collected by the NRC through the Office of Inspection and Inforcement. Inspection, violation and enforcement records are regularly compiled in a computerized file known as the "766 File". Violation data are recorded in this file according to a five level severity code defined in early 1982 [5]. A previous version of this code was based on six, rather than five, violation severity levels [6]. The violation texts listed in the file include reference to the specific requirements being violated. Most of the file records pertain to violations of QA program requirements, therefore this data source can be used, as suggested in Section 2, to evalute the level of compliance with NRC-mandated QA requirements in different plants. Summarized but still fairly detailed data and text from the 766 File are also included in the NUREG-0020 monthly reports.

4. Component Failure Rate Comparisons

The idea of comparing reliability performance data of safety-grade components with those of similar non-safety-grade components presents several problems. The identification of any such comparable sets is one of them. If for instance one is trying to identify two valve samples to compare, one should consider such factors as valve and actuator type, manufacturing materials, design pressure and temperature, pipe size and working environment, before deciding if the valves to be compared are similar engouh in design and use, so that the QA "stamp" on one of the two groups can be really considered as the driving factor behind possible differences in observed performance. The uncertainty in the possible interpretation of this kind of comparison can be made even larger by the common inconsistencies in the data bases recording criteria. For instance, a search of the LER file for the period between the beginning of 1981 and the end of 1983 showed a total of 30 events related to failures in BWR main steam relief valves (nuclear QA-grade), but only 4 related to PWR main steam relief valves (non-QA-grade). The two classes of valves are supposed to operate under similar conditions, but their designs are different and the BWR valves are sized smaller than their PWR counterparts. However, the one order of magnitude difference in the number of reported failures may have actually been influenced more by the loose reporting requirements regarding non-safety related items (like here the PWR valves) than by an underlying difference in performance due to design or QA "rating".

In some cases the problem may be originated by the difficulty of identifying the level of QA that is implemented on systems that are officially non-safety-grade. For instance, the auxiliary feedwater system (AFWS) of PWR plants is non-safety grade, but it is considered "important-to-safety", and in practice subject to most of the same maintenance practices that characterize the operation of QA-grade systems. Interestingly enough, a search of the LER data-base for events reported in 1983 on this system and on the PWR high pressure injection system (HPIS), which can be considered as a system of "similar" characteristics, showed nearly equal numbers of events, namely 133 for the AFWS and 127 for the HPIS.

Some comparisons of failure rates of safety-grade vs non-safety-grade components are made possible by using the IPRDS data compiled in Refs. [7] and [8]. Table II summarizes a comparison — performed in Ref. [7] — of the valve population of the residual heat removal (RHR) system of one plant with that of the non-safety-grade process auxiliary systems in the same plant. We also used data collected in Ref. [8] to make a comparison of the performance of safety-grade HPIS pumps with that of non-safety-grade AFWS pumps, given the existing similarities — in terms of pumping flow rates and pressure heads — between these two types of components. Table II also summarizes this latter comparison, based on the two distinct failure data categories of "catastrophic" failures (i.e., failures totally disabling the affected component) and "incipient" failures (i.e., failures detected in a non-disabling stage, but which would eventually become disabling if left unchecked). The reader should note that the valve data in Table II are for catastrophic failures only. It should also be noted that the failure on demand probabilities given in the two tables are based on a flat estimate of 12 demands/year for all the components involved. This assumption derives from the notion that most stand-by systems are tested once a month. This gives only an educated guess of the actual number of demands, however. The error introduced, especially in the case of the less important non-safety-related valves that are part of the comparison in Table II, could be substantial. Considering also that the 5-95% confidence intervals shown by the tables are quite large and vastly overlapping for the two component categories of interest (i.e., QA-grade vs non-QA-grade), it would not be prudent to draw from there a conclusion of "better performance" by one class of components over the other.

The last set of component level data that we discuss has been collected in Ref. [9]. The authors of this latter study had the benefit of access to vendor data, which provided them with a more extended statistical basis than the one afforded by the public access data bases that we have previously discussed. Ref. [9] recognizes that visual comparison of mean values and 95% confidence intervals for failure rates calculated from their data seem to indicate a better performance (i.e., lower failure rates) on the part of the QA-grade components examined in the comparison; however, it finds the significance of this indication not to be confirmed by a statistical test performed on a parameter R, which expresses the ratio of the failure rate of a given type of QA-grade component to the failure rate of a "similar" non-QA-grade type. The way in which the significance hypothesis test is formulated may actually have a non negligible influence on this outcome, and we try to show in the following that by looking at the comparison from a different perspective one may arrive at different conclusions on the interpretation and significance of the data.

Table III lists the comparison cases presented in [9]. The values $\overline{R}$, R_{025} and R_{975} are respectively the maximum likelihood point estimate, the 2.5% lower confidence bound and the 97.5% upper confidence bound derived in the reference for the R parameter defined above. The given confidence bounds are calculated by using the appropriate classical statistics formulas based on the properties of the chi-square and F distributions. The hypothesis test performed in [9] consisted in verifying whether or not the two-sided 95% confidence interval for R includes R=1. For the first five cases listed in the table it does, thus the conclusion was that R cannot be assumed to be significantly different from unity and that the failure rate of the QA-grade component type being considered is not significantly better or worse than that of the "similar" non-QA-grade component type. The last comparison case, on the other hand, does show a poorer performance by the QA-grade component.

The above hypothesis test is rather rigid in that it gives a "yes or no" answer based on an arbitrarily chosen target level of confidence (95% in the choice of Ref. [9]). In addition, the use of a two-sided confidence interval does not address the issue of interest in the most direct way. As mentioned above, the question is in whether the QA-grade components have a better or worse reliability than the non-QA-grade ones. This question may be better answered by determing the statistical confidence level at which the failure rate ratio R can be expected to be smaller than unity, or in other words, the confidence level X at which the interval (0, 1) can be assumed to include the value of R. The X values derived by both a classical and a Bayesian type of approach for each of the same cases considered by Ref. [9] are shown in the last two columns of Table III. In the classical approach we used an F-variate test as in Ref. [9], but following an inverse procedure (i.e., we fixed the confidence bounds on R and used the appropriate formula involving the F variate to determine the confidence level X). In the Bayesian calculation we interpreted R as being a random variable and took the R_{025} and R_{975} values given in Ref. [9] as the 2.5 and 97.5 percentiles of the underlying distribution. Since it is common practice by nuclear reliability specialists to assume lognormal distributions for component failure rates, and since the ratio of two lognormally distributed variables is also lognormally distributed, we assumed the distribution of R to be lognormal. It was then straightforward to use R_{025} and R_{975} to determine the parameters of the distribution and then calculate X as the cumulative probability that R be less than unity. As it can be seen, the data provide good confidence that R is less than unity in the first five comparison cases, as X is typically estimated to be in the 90-95% confidence range by both the classical and the Bayesian approaches. These indications contrast with the conclusions of Ref. [9], which accepts the hypothesis that R is not statistically distinguishable from unity for any of these cases. As we already pointed out above, the choice of conditions for the statistical hypothesis test may have here a strong impact on the conclusions that may ultimately be drawn.

5. Correlation Analysis of Plant Performance Data

As we indicated in Section 2, a way of investigating the plant-wide impact of QA practices can be pursued by carrying out empirical correlation analyses between data pertaining to plant compliance with QA requirements on one side, and plant availability on the other. To demonstrate this idea we selected eight PWR plants with about the same power ratings (i.e., in the 600-800 MWe range), then looked at the correlation factors between the QA violations recorded in the 766 File (see Sec. 3) and plant availability measures such as capacity factors or numbers of forced shutdowns in a given period of time. The time span covered in our sample analysis was from October 1980 to December 1983 included. Table IV summarizes the results by showing for which pairs of variables the correlation hypothesis was tested, what value of the correlation coefficient was obtained, and what confidence level in the existence of a relation of dependence between the two variables in the pair can accordingly be established. In the table, the "serious QA violations" column groups together the first four severity categories of the 766 File, whereas the "minor QA violations" column includes the Category V and VI violations. The analysis was performed by means of a Spearman non-parametric test [10] to avoid the restrictions regarding linear dependence which apply to linear regression analysis tests. The results of our sample analysis show a definite monotonic correspondence between plant "QA performance" and plant availability measures. These findings are very informative, although they need to be confirmed by a more refined analysis, extended to a much larger plant population and over a longer period of observation.

6. Conclusions

At the present time, a comprehensive and quantitative cost-benefit evaluation of nuclear QA programs would conceivably present serious problems, mostly due to large uncertainties in the data and difficulty in correctly interpreting the information that is available. It is possible, however, to derive from the existing data some quantitative indications of the effect of nuclear QA practices on plant component and overall plant performance.

Most of the data compilations from publicly available data bases do not contain the information needed to make meaningful comparisons of QA-grade vs non-QA-grade components. A possible exception to this could be the IPRDS data base, if further expanded to include larger component populations and improved to eliminate the existing large sources of uncertainty in component service times and demand rates.

If privately collected data were more widely accessible they could provide much useful information. Analysis of the data presented by Ref. [9] induced us to conclude that a statistically measurable better performance, in terms of reliability, could be observed in five out of six types of QA-grade components investigated there, over similar non-QA-grade components. Elaboration and analysis of publicly available plant data also seems to confirm a beneficial effect of the nuclear QA program at a global level: we observed in fact in our sample correlation study that plants with a better QA record — e.g., a lower number of QA violations — also seem to show a better overall availability — e.g., less forced shutdowns and higher capacity factors.

It should be understood that the analyses performed and presented here were carried out on a sample basis and should be extended to much larger samples before firm credit in the results can be established. Our work is mainly intended to provide some initial basis for discussion of the data utilization and analysis techniques that may be used in studies of a similar nature.

References

[1] W. D. Altman,, G. T. Ankrum et al.: "Improving Quality and the Assurance of Quality in the Design and Construction of Nuclear Power Plants", NUREG-1055, May 1984.

[2] R. Al-Ayat, S. B. Guarro, et al.: "Evaluating the Effectiveness of Quality Assurance Programs", LLNL Report, to be published.

[3] U.S. NRC/ACRS: "Review and Evaluation of the Nuclear Regulatory Commission Safety Research Program for Fiscal Years 1984 and 1985", NUREG-0963, February 1983.

[4] U.S. NRC: "Licensed Operating Reactors Status Summary Reports", NUREG-0020 (published monthly).

[5] U.S. Federal Register, Vol. 47, No. 46, March 1982.

[6] U.S. Federal Register, Vol. 45, No. 196, October 1980.

[7] R. J. Borkowski, W. J. Kahl, et al.: "The In-Plant Reliability Data Base for Nuclear Plant Components: Interim Report - The Valve Component", NUREG/CR-3154, December 1983.

[8] J. P. Drago, R. J. Borkowski, et al.: "The In-Plant Reliability Data Base for Nuclear Plant Components: Interim Data Report - The Pump Component", NUREG/CR-2886, December 1982.

[9] G. K. Roberts, S. J. Sarver: "A Study of the Role of Quality Assurance in Nuclear Power Component Reliability", Westinghouse Electric Corporation, September 1984.

[10] W. J. Conover: "Practical Nonparametric Statistics", J. Wiley & Sons, 1980.

Acknowledgments

This work was partially supported by the U.S. NRC under a memorandum of understanding with the Department of Energy.

Table I: Important aspects of a nuclear QA program

QA PROGRAM		
Activities	Directly Affected Costs	Impact Areas
Quality control	Program Management	Public Safety
QC procedures enforcement	Schd. maintenance & repairs	Component & plant availability
Procedure & activity documentation		Public confidence
		Traceability

Table II: Reliability comparison of QA vs. non-QA components

	QA valves	non-QA valves	QA pumps			non-QA pumps			
Component Characterizations / Failure Parameters	R1	P1	H1MC	H1MS	H2MX	A1D	A1T	A2M	A2T
Q_{CL}	1.4	0.96	11	0.43	6.2	0.86	0.86	-	-
Q_{CM}	2.9	1.6	33	8.3	35.	17.	17.	-	-
Q_{CU}	5.5	2.5	76	40.	110.	79.	79.	75	140
Q_{IL}	-	-	100	70	57	120	55	22	75
Q_{IM}	-	-	160	120	120	200	120	78	210
Q_{IU}	-	-	230	180	230	320	220	190	420

Parameter symbols and suffixes:

Q = failure frequency x 10^3 (per demand)

C = catastrophic failure   I = incipient failure

L = 5% lower bound   M = mean   U = 95% upper bound

Component characterizations:

a) System:

A = auxiliary feedwater   H = high pressure injection

R = residual heat removal   P = process auxiliaries

b) Plant:

1 = Plant "One"   2 = Plant "Two"

c) Pump drive:

D = diesel   M = motor   T = turbine

d) Pump type:

C = charging   S = safety injection   X = charging/high head injection

Table III: Hypothesis testing on component performance comparisons.

COMPARISON CASE (QA-grade component is listed first in each pair)	R_{025}	$\overline{R}$	R_{975}	X = confidence level for hypothesis R < 1: classical	assuming lognormal distr. for R
1. Reactor Coolant Pump Motors vs. Condensate Pump Motors	0.22	0.66	2.02	72%	78%
2. Reactor Coolant Pump Motors vs. Main Feedwater Pump Motors	0.13	0.43	1.50	90%	91%
3. Residual Heat Removal Pump Motors vs. Heater Drain Pump Motors	0.07	0.36	1.17	95%	96%
4. Reactor Coolant Pump Canned Motors vs. Boiler Water Circulation Pump Canned Motors	0.005	0.19	1.23	95%	97%
5. Westinghouse Mod. DS Circuit Breakers (Nuclear vs. Non-Nuclear)	0.14	0.51	1.32	89%	93%
6. Residual Heat Removal Heat Exchangers vs. Low Pressure Feedwater Heaters	1.16	1.89	2.99	0.4%	0.4%

Table IV: Non parametric correlation test of plant availability and QA performance data.

	Availability Indicators			QA Performance Indicators			Correlation Analysis Results	
Test Cases	Forced Shutdowns	Forced Shutdowns and Power Reductions	Capacity Factor	Serious Violations	Minor Violations	Total Violations	Spearman Correlation Factor	Confidence Level in Existence of Correlation
1	X			X			0.815	99%
2	X				X		0.637	95%
3	X					X	0.744	97.5%
4		X		X			0.649	95%
5		X			X		0.625	95%
6		X				X	0.571	92.5%
7			X	X			-0.565	92.5%
8			X		X		-0.601	95%
9			X			X	-0.609	97.5%