INL/EXT-20-58327

Evaluation of Weakly Informed Priors for FLEX Data

Cindy Gentillon
Corwin L. Atwood
Andrea L. Mack
Zhegang Ma

May 2020

INL/EXT-20-58327

Evaluation of Weakly Informed Priors for FLEX Data

Cindy Gentillon[1], Corwin L. Atwood[2], Andrea L. Mack[1], Zhegang Ma[1]

Completed May 2020

[1]Idaho National Laboratory
Nuclear Safety and Regulatory Research Division
Idaho Falls, Idaho 83415

[2]Statwood Consulting
2905 Covington Road
Silver Spring, MD 20910

http://www.inl.gov

Prepared for the
Division of Risk Analysis
Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission
NRC Agreement Number 31310019N0006
Task Order Number 31310019F0022

ABSTRACT

On behalf of the U.S. Nuclear Regulatory Commission, Idaho National Laboratory (INL) reviewed the development and use of weakly informed prior (WIP) probability distributions to obtain industry-wide failure probability and rate distributions for portable FLEX equipment, as documented in the Pressurized Water Reactor Owner's Group (PWROG) draft report, *FLEX Equipment Data Collection and Analysis* (PWROG-18043-P Revision 0) by M.M. Degonish. The review includes a description of the methods used by the PWROG analysts. It also includes an implementation of those methods for comparison purposes, as well as an implementation of two constrained noninformative (CN) distributions for further WIP comparisons.

The report documents several issues that the review identified in the development of WIP distributions. They relate to the quality and representativeness of current portable FLEX component data, the choice of informed distributions for related installed components, and the lack of verification for factors chosen to weaken the information from the installed components. Two errors occurred in the computations: the posterior distribution of the WIP was used to modify the WIP itself, resulting in a lack of independence between the WIP and the data; and an error was made in the conversion of WIP mean and variance values into beta distributions.

Due to these issues, INL reviewers believe the WIP distributions presented in the report to be deficient for use in risk assessments at the present time.

EXECUTIVE SUMMARY

On behalf of the U.S. Nuclear Regulatory Commission, Idaho National Laboratory (INL) reviewed the development and use of weakly informed prior (WIP) distributions to obtain industry-wide failure probability and rate distributions, as documented in Pressurized Water Reactor Owner's Group (PWROG) draft report PWROG-18043-P (Ref. [ES-1]). The WIP distributions were for portable FLEX components that increase flexibility in nuclear power plants (NPPs) when responding to beyond-design-basis external events. The distributions are intended to characterize the uncertainty and variability in associated failure probabilities and rates across the NPP industry, and are for use in probabilistic risk assessment.

Most observations found in the review relate to the five steps used by the PWROG to create recommended WIP distributions: (1) obtain data, (2) select relevant industry information about installed components to use for informed distributions, (3) determine factors to make the distributions only weakly informed and to account for uncertainty in applying data from installed equipment to the portable FLEX equipment, (4) fit the resulting mean and variance values to beta distributions for failure to start (FTS) and gamma distributions for failure to run (FTR), and (5) perform Bayesian updates using the FLEX data for each component. The resulting posterior distributions are reported in Table 6-1 of the PWROG report. Observations regarding each of these steps are listed below.

Data Observations

- Data for several of the 16 components studied are very limited. For example, only two actual medium-voltage diesel-driven positive displacement pumps from a single site provided data for the study.
- Some FLEX data are over a decade old. Components in that category are over-represented in the data. Such imbalance in the data is likely to result in biased outputs.
- Data collected prior to implementation of the FLEX guidelines initially issued in 2012 are suspect, since testing and reporting might not be as consistent and thorough.

Selection of Data for Informed Distributions

- Though none of the industry generic data were differentiated by operating conditions, the FLEX data were. For example, FLEX diesel generators were split into three groups based on output voltage, whereas the industry data make no such distinction. More detailed industry data might be found, or the FLEX data might be pooled (after testing whether the data reject the hypothesis of homogeneity among these groups).
- Data for installed diesel-driven centrifugal pumps were applied to make distributions for positive displacement pumps.
- The application of data for failure modes "FTR<1H" and "FTR>1H" needs more study.
- Some use of industry data might be validated by comparing industry mean values with FLEX data from categories that contain more FLEX data.

Adjustment Factors

- Without verifiable justification, "4" was selected as the mean and standard deviation multiplier for each component to both raise and widen the distributions and make WIPs.
- In Bayesian data analysis, the data used to update a prior must be independent of that prior. In PWROG WIP development, the variance multiplier (E_M) is tuned by evaluating the performance of the posterior distribution. As E_M changes, the WIP changes. Such dependence is unacceptable.
- Even if selection of E_M was based on the prior distribution alone, the criterion (based on "range factors") guiding the selection is unachievable for some industry distributions.

Fitting Beta and Gamma Distributions

- There was an error in the calculations for beta distributions. As a result, the variances of the four WIP FTS distributions in the report do not satisfy the condition of being {4*4*E_M} times the industry distribution variances. In the review, the WIP distributions were recalculated.

Additional Observations

- For the components in the PWROG report with WIP distributions, the (corrected, if necessary) WIP distributions were compared with three alternative distributions: the Jeffreys noninformative prior (JN) distribution and two constrained noninformative (CN) prior distributions. All CN distributions are designed as wide distributions subject to a constraint on the mean value. One of the CN distributions matched the WIP mean (4* the industry value), and the other had a mean equal to ten times the industry value. All three alternative distributions had wider 5%-95% intervals than the WIP distributions for FTR. For FTS, the WIP distribution widths were similar to the CN prior distribution widths with the same prior mean. The FTS WIP alphas were sometimes very small. The JN distributions, using only FLEX data, were the most conservative in all cases.
- To observe the effects of changing the operating experience boundary between the components modeled with WIP and those modeled with JN, INL reviewers applied the JN and PWROG WIP (with corrected FTS calculations) methods to all components and failure modes not modeled using empirical Bayes (EB). The posterior JN distributions had higher means and 95th percentiles (with just one exception), especially when demands or operating hours were low.
- The review team benefitted from additional material received from the authors of the PWROG report and discussions with them via conference calls. Inclusion of much of this material in the report would be beneficial, particularly in regard to the values of the variance factors and the full details about the installed equipment distributions used as initial sources.

Final Remarks

The INL reviewers believe the WIP distributions presented in the report to be deficient for use in risk assessments because of the following issues:

- Data used
  - Not enough components with current FLEX data to be representative of the nuclear plant industry
  - No attempt to validate the use of the selected generic industry data for installed components
- Analysis method
  - No attempt to validate the scaling factor of 4
  - Use of the posterior distribution to adjust the WIP
  - The small calculational error affecting the FTS WIPs.

Reference

[ES-1] Degonish, M.M., 2020, *FLEX Equipment Data Collection and Analysis*, PWROG-1803-P, Rev. 0.

CONTENTS

FIGURES

TABLES

ACRONYMS

CN	constrained noninformative
EB	empirical Bayes
EF	error factor
FTR	fail to run
FTS	fail to start
INL	Idaho National Laboratory
JN	Jeffreys noninformative
NPP	nuclear power plant
PWROG	Pressurized Water Reactor Owner’s Group
RF	range factor
UF	upper fence
WIP	weakly informed prior

Evaluation of Weakly Informed Priors for FLEX Data

1. BACKGROUND

The Pressurized Water Reactor Owner's Group (PWROG) uses weakly informed priors (WIPs) for certain industry distributions for failure to start (FTS) and failure to run (FTR) for portable equipment used to mitigate beyond-design-basis external events in nuclear power plants (NPPs). FLEX equipment is intended to increase plant flexibility in responding to such events. This paper contains a brief description of the WIP method proposed in PWROG-18043-P Revision 0 (Ref. [1]) for FLEX components having little operational data. It then evaluates the suitability of the method.

The body of this paper evaluates the WIP and its implementation and compares it with possible alternatives. Appendices provide additional details on the following topics:

Appendix A. Theoretical comparison of WIP and the constrained noninformative (CN) prior

Appendix B. Use of range and error factors

Appendix C. Evaluation of the cutoff values in places where the PWROG report switches from one estimation method to another

Appendix D. Comments on Appendix A of Ref. [1], which contains a pooling analysis

Appendix E. Review of the study of outliers, given in Appendix B of Ref. [1].

The primary purpose of this report is to evaluate the use of WIP. The next two sections provide background information about the FLEX data set and the process through which the data are used to derive industry uncertainty distributions for the portable FLEX equipment. The bulk of the report is in the evaluation and comparisons that follow. Other comments on the rest of Ref. [1] are occasionally made, especially in the Appendices.

1.1 FLEX Data

The data for the new, portable FLEX components are summarized in Table 1. The data were pulled from several tables, in addition to Appendix B in the PRWOG report. As can be seen, the data are divided into seven types of components typically operating under various conditions that differ in terms of voltage, pressure, flow rates, etc. In all, there are 16 classes of components/operating conditions. Each class exhibits two failure modes: FTS and FTR. The parameters to be estimated for these two modes are the probability of failure on demand (p) and the rate of failures per hour (λ).

The amount of data varies greatly among the classes, from thousands of demands or running hours with as many as a few dozen failures, down to merely a few demands or running hours and only one potential failure. Ref. [1] pp. 5-6 and 5-7 states the ways that failure parameters are estimated: use a WIP for classes with fewer than 50 demands or 100 hours of operating experience; otherwise, use the empirical Bayes (EB) method or the Jeffreys noninformative prior (JN), depending on whether a statistical test does or does not reject the hypothesis that the plant data can be pooled.

Table 1. PWROG FLEX data summary [a].

Portable Comp.	Driver	Op. Cond.	# sites with comp.	# of comp	FTS (beta distributions) # F	# Dem	Mean	α	Meth.	FTR (gamma distributions) # F	Run Hours	Mean	α	Meth.
Generator	Diesel	Installed	—	—	—	—	2.88E-03	—	—	—	—	1.52E-03	—	—
		High V	4	9	1	176	8.47E-03	1.5	JN	1	175.0	8.57E-03	1.5	JN
		Med. V	50	167	67	2346	3.13E-02	1.68	EB	12	4086.4	3.51E-03	0.47	EB
		Low V	1	6	2	27	2.40E-02	3.24	WIP	0	8.9	5.83E-03	1.26	WIP
	Com-bustion Turbine	Installed	—	—	—	—	5.12E-02	—	—	—	—	8.49E-03	—	—
		High V	4	8	1	146	1.02E-02	1.5	JN	1	124.7	1.20E-02	1.5	JN
		Med. V	5	13	6	158	4.09E-02	6.5	JN	1	119.3	1.26E-02	1.5	JN
Centrif. Pump	Diesel	Installed	—	—	—	—	2.17E-03	—	—	—	—	9.80E-04	—	—
		High Pr	7	20	8	223	1.26E-01	0.2	EB	3	306.3	1.14E-02	3.5	JN
		Med. Pr	37	129	24	1659	1.65E-02	1.04	EB	12	2468.7	6.15E-03	0.94	EB
		Low Pr	33	99	40	1666	2.36E-02	0.997	EB	9	1424.3	9.31E-03	0.25	EB
Positive Displ. Pump	Diesel	Installed	—	—	—	—	2.17E-03	—	—	—	—	9.80E-04	—	—
		High Pr	6	10	0	133	3.73E-03	0.5	JN	0	97.5	2.87E-03	1.05	WIP
		Med. Pr	1	2	0	22	7.63E-03	1.39	WIP	0	17.0	3.69E-03	1.05	WIP
		Low Pr	0	0	0	0	8.68E-03	1.39	WIP	0	0.0	3.92E-03	1.05	WIP
	Motor	Installed	—	—	—	—	7.49E-04	—	—	—	—	1.22E-04	—	—
		High Pr	15	44	3	507	6.89E-03	3.5	JN	2	254.9	9.81E-03	2.5	JN
		Med. Pr	3	26	2	78	3.16E-02	2.5	JN	1	35.4	1.05E-03	1.84	WIP
		Low Pr	1	3	0	57	8.63E-03	0.5	JN	0	27.0	4.80E-04	0.84	WIP
Air Com-pressor	Diesel	Installed	—	—	—	—	8.24E-03	—	—	—	—	2.36E-04	—	—
		—	14	46	6	293	2.21E-02	6.5	JN	4	310.3	1.52E-02	0.17	EB
	Motor	Installed	—	—	—	—	4.16E-03	—	—	—	—	1.35E-05	—	—
		—	4	7	1	34	2.09E-02	2.12	WIP	0	19.0	5.36E-05	1.13	WIP

a. Shaded-green rows show the nominal mean values obtained from external sources for installed components that might behave similarly to the associated FLEX components. Blue-shaded cells flag instances in which the WIP was used.

1.2 Summary of PWROG WIP Process

As described in Ref. [1], a WIP is selected for estimating FTS probabilities for components with less than 50 demands, and for estimating FTR rates for components with less than 100 operating hours. For these portable components, the information carried in actual operating data is deemed insufficient. The recommended distributions for p or λ in these cases are obtained by examining the literature to find established industry distributions for similar installed components. These data form the basis for *informed* prior distributions.

Because the new FLEX components may be less reliable than the installed components or involve extra difficulty in their use, the nominal values obtained above are adjusted upward. Ref. [1] constructs WIP distributions by multiplying the means and standard deviations of the informed priors by factors that both widen the distributions and increase their means. This adjustment process makes the priors less informative and more consistent with the limited data. The PWROG calls such widened distributions "WIPs." The priors are conjugate priors, meaning they are gamma distributions for rates and beta distributions for probabilities. The Bayesian update process for these distributions is simple. The equations are on p. 5-7 in the PWROG report.

The scaling factors are intended to account for various uncertainties associated with the portable equipment. There are four possible factors, described on pp. 5-10 and 5-11 in Ref. [1]. Additional factors are described in Ref. [2].

The combined effect of the Ref. [1] scaling factors is that the mean values and standard deviations are each multiplied by 4, regardless of the component or failure mode. Whether these particular selections are best is ultimately a matter of engineering judgement. The report does not describe any attempts to validate or verify this particular multiplier.

One other factor is involved in the process. It increases the variance of the distribution without changing the mean. The variance factor is determined by considering range factors (RFs). Like the lognormal distribution error factor (EF), an RF is defined as the square root of the ratio of the 95th percentile of a distribution and its 5th percentile. If the RF of the initial prior distribution (for installed equipment) is less than 5, an additional factor (variance factor = E_M) is multiplied by the source variance. An RF of 5–10 (or even as high as 15) is deemed acceptable for a WIP distribution. These topics are discussed further in Appendix B.

If the posterior distribution has a small RF, the E_M term is revisited and increased. This process is repeated until the posterior distribution RF falls somewhere between 5 and 10, if possible.

Finally, a new gamma (for rates) or beta (for probabilities) distribution is identified by fitting its parameters so that the distribution has the desired mean and variance. The equations for this process are on p. 5-12 of Ref. [1].

2. CRITIQUE OF WIP CONSTRUCTION

The fundamental process in a traditional Bayesian analysis is to use an informed prior distribution based on phenomena that resemble the situation under study, then update that prior with directly relevant observed data. The process requires that the prior and data be compatible. If the prior assigns low probability to an interval that contains the failure rate or probability estimates from the data, it is not compatible. The resulting posterior distribution will be close to the prior, despite the data. If different mechanisms govern the prior distribution and new data, a broad, less-informative prior distribution will more likely be useful.

The idea of applying factors to the informed prior to obtain a useful WIP is not unique. For example, in human error analysis, performance-shaping factors are applied to probability mean values published for situations similar to that under analysis.

In the sections below, and later in the Appendices, the evaluation of the PWROG WIP considers each of the four steps involved in using a WIP to obtain a posterior distribution for a FLEX component: collecting FLEX operational data, selecting informed distributions, selecting the factors that make these distributions "weakly informed," and fitting beta or gamma distributions to the adjusted means and variances. There is also an evaluation of the cutoff used to determine whether a WIP distribution is appropriate for a component, or whether the Jeffreys noninformative prior would be better.

2.1 Statistical Perspective on FLEX Operational Data

Review of the operational data is primarily an engineering pursuit. However, from a statistical point of view, questions arise about whether the data are representative of the NPP industry as a whole. When the data come from a limited number of sites, or, even worse, a limited number of actual components, the results are likely biased by whether those sites or components are typical. Table 1 lists the number of sites with components, along with the number of components for which data are available.

A similar concern relates to the use of data before 2012. A few components have histories going back into the 2000's. These have an unbalanced effect on the data, due to the disparity in the volume of their data compared with that of other components. This is explored further in Appendix E, in which some of the possible outliers appear to be sites with extremely long histories compared to other sites.

The data need to be representative of the industry as a whole, in addition to being both complete and accurate. This means that, for the components being tracked, all failures, demands, and operating hours must be reported. Further, the demands must stress the components in the same way that actual demands would. Because FLEX components are offline, they receive no real demands associated with heat removal or other types of plant protection. Testing must adequately simulate the conditions the component would face if brought into use.

2.2 Evaluation of Process of Selecting Informed Distributions

Determination of a WIP distribution for a portable FLEX component starts with selecting an informed distribution that describes the performance of similar installed NPP components. Ref. [3] provides a robust dataset suitable for use in risk assessments, covering over 200 NPP component/failure mode combinations. A spreadsheet (Ref. [4]) provides updated estimates of those failure rates and probabilities, using data through 2015. However, there are issues associated with selecting estimates from this source when it comes to the 16 FLEX components listed in Table 1.

The first problem is that the source data do not provide the level of detail about operating conditions that is needed by the FLEX list. In every case in the FLEX data table that provides estimates for installed equipment, the specified operating conditions are not addressed by Refs. [3] or [4]. If it is necessary to have separate estimates based on voltage, pressure, or flow rate, separate estimates for the installed data are needed. Otherwise, one might expect additional uncertainty factors to be employed in setting up the prior distributions.

Alternatively, the operating condition breakdowns might be omitted.

The same concern applies to the pump-type associated with diesel-driven pumps. The source data does not distinguish between centrifugal pumps and positive displacement pumps (PDPs) when the driver is a diesel. Perhaps these categories should be combined. Or, alternative sources of industry data might be found. Otherwise, additional uncertainty factors may need to be used when information is present for which pump designs are more reliable.

For motor-driven pumps, the opposite is true. The source provides separate estimates for both standby centrifugal and PDP motor-driven pumps. FTS estimates for these components are not needed for the FLEX study, since the FTS data in each operating condition category has enough demands. However, motor-driven PDP estimates for FTR exist and should probably be used.

The last issue concerning the selection of mean values for informed prior distributions is that, for several components, Refs. [3] and [4] split FTR into two categories. For diesel generators, the split is between failure to load and run (FTLR) and FTR. For others, the split is between FTR during the first hour of operation (FTR<1H) and FTR afterwards (FTR>1H). In the study leading to Ref. [2], the rate of failure occurrence was found to be higher in the first of the two periods. For diesel generators, the FLEX informed prior distribution was based on FTR rather than FTLR (which shows a little over double the occurrence rate, though this may be for generators with more extensive loadings than seen by the FLEX generators). In any case, these selections need to be justified.

Overall, the report provides insufficient information for a complete evaluation. It does not provide the alpha parameter or standard deviation of the selected distributions, nor does it specify the particular source for each selection. (This information, however, was later supplied to the INL review team).

2.3 Critique of Process to Obtain Adjusting Factors

Two types of factors were applied to the initial distributions to make them weakly informed and better suited to the FLEX components: one that increases the scale of the input distributions, and a separate factor to enlarge the variance. Issues concerning these factors are individually discussed below. Ultimately, the choices of values for the scaling factor, desired range factors, and E_M factors were based on engineering judgment, without further justification.

2.3.1 Critique of the Scaling Factor

The FLEX data differ from the installed components represented in the 2015 Update, from the perspective of different operating conditions as well as different component types. From a statistical point of view, factors in addition to those listed in the report are needed to account for these differences. Also, the report does not go into much detail about why the location factor is 1.0. Perhaps the storage environment is comparable to the environment in which the component will be used. The process of getting the component from its storage location to where it is needed must be modeled in other aspects of a probabilistic risk assessment.

It is possible to back-calculate and observe the mean and standard deviation of the WIP, even though this information is not provided in the report. The information was transmitted privately for use in this review, but the INL reviewers believe it should be readily available in the report.

The PWROG chose to use 4 as the total multiplier for all distributions. The accuracy of this multiplier can be checked in cases with ample FLEX data. For example, consider turbine-driven generators with medium voltage and FTS. There were six failures in 158 demands, and a posterior distribution based on the JN prior was found. Compare the 5^{th} and 95^{th} percentiles with 4 times the installed mean to see whether the adjusted industry value and the data seem compatible. With uncertainty information for the industry value, one can even compare the two intervals: installed vs. FLEX data. Such comparisons for all cases with adequate FLEX data can confirm the factor of 4 is in the right ballpark—or in some cases, maybe not.

This technique does not violate the proscription against using the data to construct the prior. The data from components with a moderate or large number of failures are used to estimate the adjustment factor. This factor (4 in Ref. [1]) is then used in the analysis of components with very few failures. This follows the standard process of using similar data to construct a prior distribution, but not the very same data to be used in the Bayes update.

2.3.2 Critique of the Variance Factor

During a teleconference discussion [5], the PWROG authors stated that the variance factor (E_M) was derived through trial-and-error to give a "reasonable" posterior distribution, one without too small of a variance. They were trying for a posterior RF of 5–10, which they thought realistically expressed their uncertainty. The reviewers are skeptical of the processing for this factor for the following reasons:

- In the Bayesian process, the data have a mathematically specified influence. As more data are collected, the posterior distribution narrows. If the quantity of data increases, the RF of the posterior distribution should decrease. Trying to manipulate the posterior distribution to keep its RF above a certain number obviates the value in collecting more and more data. The posterior distribution must remain free to respond to the data.
- Conversely, a prior distribution must be *prior*—the data should not be used in constructing the prior distribution, neither directly nor through the posterior distribution. If a posterior distribution is not to someone's liking, one correct response is to obtain additional data. An incorrect response is to go back and change the prior.
- In addition, skepticism exists because of the innate human inability to estimate uncertainty. An influential article by Tversky and Kahneman [6] claims that, unless a special effort is made, people tend to anchor on a point estimate of an unknown quantity, then insufficiently adjust to account for their uncertainty, resulting in an interval estimate that is too short. This claim has been the subject of many articles. For example, Block and Harper [7] describe six experiments that generally confirm the earlier claim, though details of the anchoring and estimation process can influence the overconfidence. In the final paragraph of their article, they state:

  > Because overconfidence in estimating unfamiliar quantities is especially pervasive at higher confidence levels (e.g., 90%), even experts may need to take steps to decrease such inappropriate confidence.

  The PWROG staff evidently did not take such steps. Their process was to start with a mean that is 4 times that of industry installed components, then to adjust the E_M factor for the variance until a modest RF was obtained. This is almost exactly the process described by Tversky and Kahneman. Instead of showing that the information in the data along with the multiplication by 4 corresponds to the chosen RF, they simply assert that this RF seems right. According to Refs. [6] and [7], the PWROG estimate is probably overconfident, with an RF that is too small.
- Finally, the choice of E_M is not objectively defined. It depends on undocumented choices made by the analyst. If the reviewers or Nuclear Regulatory Commission staff were to independently try to use the WIP method for a dataset other than the eleven identified as WIP in Table 1, there is no guarantee that the exact same E_M selections would be made as those chosen by the PWROG staff.

The process used by the PWROG to determine E_M is not always successful. For three of the eleven PWROG WIP distributions, the posterior RF was less than four. The possibility of the method working with a different definition of RF (such as the 95th percentile divided by the 50th) was considered by the reviewers. Looking at the ratio of the 95th and 50th percentiles is more appropriate for skewed distributions and where upper limits of failure rates/probabilities are of most concern. Because the RF is similar to an error factor (EF), Appendix B contains a discussion of different ways to define EFs, along with a table of EF values that would be considered appropriate for various circumstances. However, even with the second RF definition, for some components, no E_M value would allow the RF to fall between five and ten.

In addition to these conceptual issues, there was a serious problem in execution, as discussed in the next section.

2.4 Error in Fitting Beta Distributions

The PWROG report uses the scale factor and E_M values in a mathematically correct way for gamma distributions and FTR. However, there is an error in the application for beta distributions. One of the formulas on p. 5-12 was incorrectly evaluated. As a result, all four WIP distributions for FTS are incorrect. The stated alpha and beta lead to beta distributions with variances approximately 6–200 times the source variances, whereas the factors would lead one to expect the ratio to be 16* E_M. For the four FTS distributions, the factors were 30, 30, 350, and 35, respectively, for FTS for diesel-driven positive displacement pumps, both low voltage and medium voltage; low-voltage diesel generators; and motor-driven air compressors.

It is not enough to simply correct the calculation formula. For FTS, the calculations were corrected and variance factors identified using the PWROG methodology. The new prior alpha parameter was much lower than before. Among the posterior RFs for the four FTS WIPs in the PWROG report, two were bounded such that no variance factor could make the RF less than 10. For another entry, the opposite occurred: no variance factor could make the RF 5 or greater. When the posterior RF is bounded away from the [5, 10] interval, no E_M value will bring it into the desired range. In re-evaluating the FTS WIP distributions, INL reviewers selected low variance factors that put the RFs either between 5 and 10 (but nearer to 5) or near their limiting value.

For the four distributions analyzed by both INL and the PWROG, the variance factors changed from the PWROG values cited above for {PDP_D_LP, PDP_D_MP, GEN_D_LV, and CPR_M} to the INL values {1, 1, 35, and 3}, respectively.

2.5 Evaluation of Cutoffs for Using WIP

In the PWROG methodology, the cutoff between using WIP and updating the JN prior to obtain an uncertainty distribution for a component is based on the amount of operational data. Specifically, the WIP is selected if less than 50 demands or less than 100 run hours exist in the data. Difficulties discussed in the previous sections may render moot the question of the adequacy of these cutoffs; but the evaluation was performed, and a summary of the results follow. The details are found in Appendix C.

For both FTS and FTR, the WIP posterior means were generally lower than the corresponding JN values. The only exception occurred for FTR among turbine-driven generators with medium voltage: 119 run hours and one failure. This is the only component class with higher estimates with the WIP distribution than with the JN. Therefore, apart from that one exception, raising the cutoffs generates lower (WIP) estimates, and lowering them generates higher JN estimates. The graphs show both sets of estimates coming together as the exposure time or demands increase. With more time, the data have greater influence on both types of priors, pulling them to a common data-influenced outcome.

3. ALTERNATIVES TO USING WIP DISTRIBUTIONS

In the PWROG report (Ref. [1]), three methods were used to obtain uncertainty distributions: the WIP, JN, and EB methods. All these methods produce beta distributions for probabilities and gamma distributions for rates. The EB method is not an alternative to the WIP distribution, however, because the maximum likelihood estimation process diverges when the data are insufficient.

The JN method uses only the FLEX data (failures, f; demands, d; or operating hours, T). It produces a fairly broad distribution when failures are few.

Another Bayesian approach used to produce uncertainty distributions is the CN distribution. Here, the prior distribution is a relatively flat distribution containing little information, except that it is constrained to have a specified mean value. The alpha parameter of the CN distribution is always 0.5 for gamma distributions and near 0.5 for beta ones. This distribution is described in detail in Appendix A. Two values were considered for the specified mean:

- The prior mean used by the PWROG for the WIP, which equals 4 times the industry installed component mean selected from Ref. [4]
- Use 10 times the industry installed component mean selected from Ref. [4]

In the plots below, posterior means and 5th/95th percentiles are displayed for the two CN distributions, in addition to the WIP means and percentiles and the JN means and percentiles. Figure 1 provides comparisons for the four components with WIP distributions for FTS. The distributions are not the ones presented in the PWROG document, due to the aforementioned error. Instead, the WIP process was followed for the four components using the correct formulas. The new beta distribution parameters and E_M values for the components are listed in Appendix C.

In Figure 1, component labels are in the gray boxes on the left side of the plot. The first three letters describe the component type, as follows:

CPR — Compressor

GEN — Generator

PDP — Positive Displacement Pump

The fourth character describes the driver: D (diesel) or M (motor). The third label, if given, describes the operating conditions. Here, LV is low voltage, MP is medium pressure/medium flow rate, and LP is low pressure/high flow rate. Also, in Figure 1, the WIP means and percentiles are labeled “WIP,” the two CN distribution outputs are labeled “Installed*4_CN” and “Installed*10_CN,” and the Jeffreys output is labeled “JN”.

Figure 2 is similar to Figure 1, except that it shows results for the seven components with WIP distributions for FTR. The WIP distributions are from the PWROG report. The component types and drivers are the same as those for the FTS WIP plot. An additional operating condition code, “HP,” stands for high pressure/low flow rate for pumps. Acronyms for the distribution methods are the same as in Figure 1.

In each figure, the components are sorted so the ones at the top of the figures are those with the most demands or run hours, and the components at the bottom are those with the least operational data.

Two features in Figure 2 deserve further explanation. First, there is no JN plot for the component at the bottom of the figure (PDP_D_LP). The diesel-driven positive displacement pumps with low

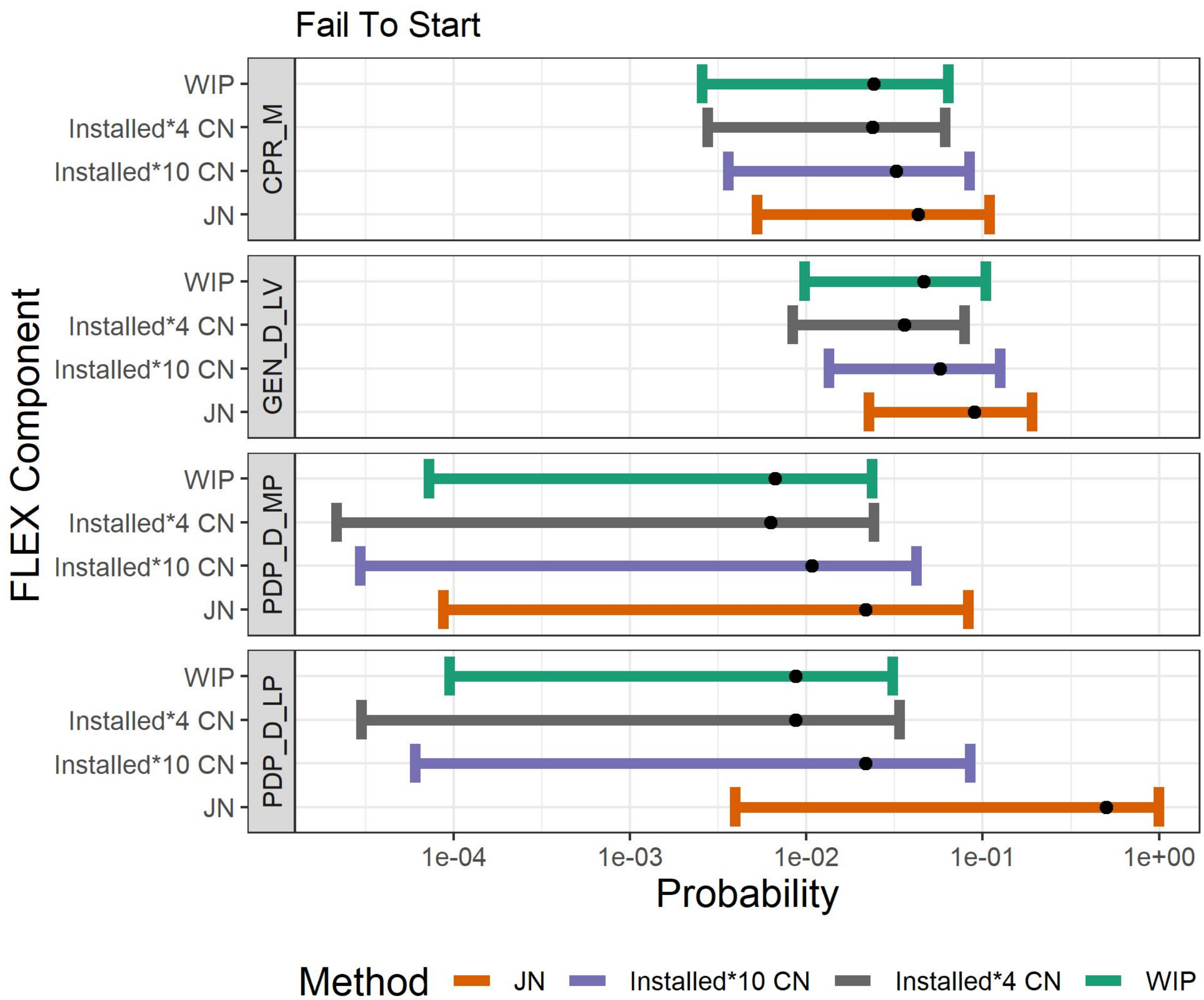

Figure 1. Posterior distribution 5th percentiles, mean values, and 95th percentiles for the probability of failure to start for components with WIP distributions.

pressure/high flow rates had absolutely no data (no operating hours, no demands, no failures), though they are listed in the PWROG study. For the JN method for gamma distributions, the data mean is (f+0.5)/T, so run time must be greater than 0 to evaluate the posterior mean.

The second anomaly in Figure 2 is the lower WIP and CN distribution bands for motor-driven compressors (the middle set of bands in this plot). The values are low because the FTR industry value for installed compressors is an order of magnitude lower than the other values used in the FLEX component study.

Both figures show similar observations:

- The means and bounds generally have a smaller spread for the components at the top of the figures, and a larger spread for those at the bottom, reflecting the impact of more operational experience. The narrowest bands, however, are for the components that had failures. In Figure 1, the top component had one failure and the second had two. The other components had no failures. In Figure 2, no components had failures, except for the one in the second band (PDP_M_MP) that had one failure.

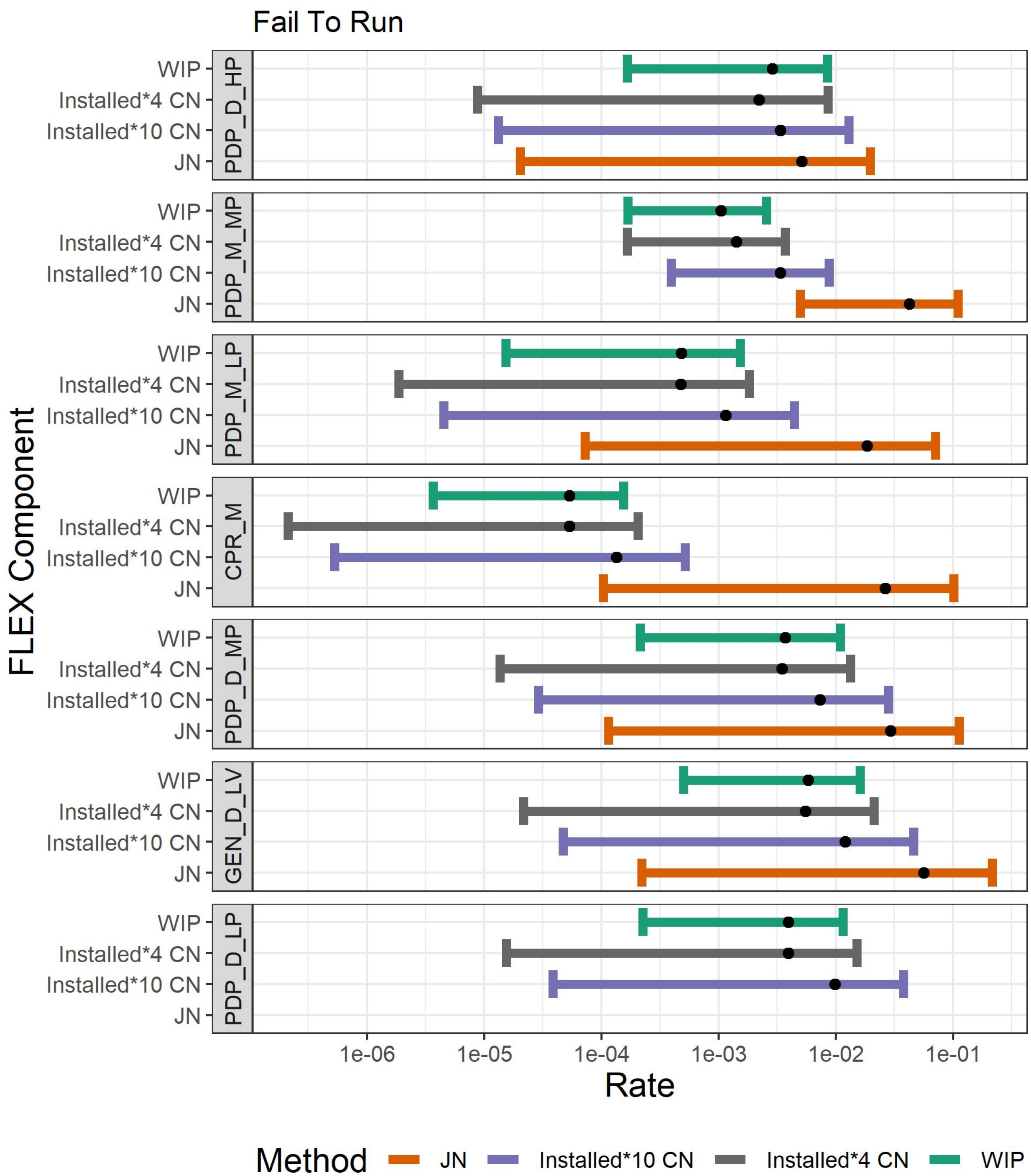

Figure 2. Posterior distribution 5th percentiles, mean values, and 95th percentiles for the rate of failure to run for components with WIP distributions.

- In virtually every case, the JN method has the highest mean and the highest 95th percentile. Among the methods considered, it is the only one not influenced by the industry installed equipment mean values.

- The two CN distributions have identical spreads for their bounds because they have the same alpha parameter for FTR rates, as well as similar alpha parameters for FTS probabilities. As expected, the Installed*10_CN bounds and mean are to the right (higher probabilities and rates) than the Installed*4_CN values.
- The WIP distribution posterior mean values are close to the Installed*4_CN posterior mean values. Both distributions have the same prior mean.
- In every case in Figure 2 (FTR distributions), the WIP posterior distributions appear narrower than the other distributions. This is not so apparent in Figure 1 (FTS), which reflects the changes that occurred when the algorithm for finding beta distributions with specified means and variances was corrected. One effect of the change was that the prior alpha parameters became much lower. The change widens the prior distribution, thus somewhat widening the posterior FTS distributions, as well.

Various alternatives exist for developing uncertainty distributions for the WIP components. For example, the FLEX medium-voltage diesel data could be used to establish a prior distribution for both the low-voltage and high-voltage diesel generators. The effect of using positive displacement pump data from installed components could be examined. Similarly, the use of other failure modes from Ref. [4] (FTLR, FTR<1H) could be observed. The effect of pooling the data across operating conditions could be investigated, though this would require site-level data. A sensitivity study could be performed, excluding data prior to the plant's implementation of the FLEX program. Factors other than four could be considered for the WIP means, and other variance-scaling factors could be considered, as well. These alternatives are beyond the scope of the present report. The reviewers recommend exploring other options before making a final decision on the distributions. They do not support the WIP distribution as implemented in the PWROG report.

4. ADDITIONAL TOPICS

Additional topics include a review of the appendices from the report. The fact that the issues of pooling and outliers were considered is commendable. The reviewers also make some general observations about the level of detail supplied in the report.

4.1 Pooling Evaluation Summary

Appendix B of Ref. [1] contains an evaluation of whether data may be pooled across sites for each of the report's 32 portable FLEX component distributions (16 FTS and 16 FTR). A chi-square statistic was evaluated for every case in which data came from more than one site. The possibility that typical chi-squared distribution critical values might not be appropriate is addressed but not acted upon. In nearly all the evaluations, the data were too sparse (an expected value of less than 0.5 failures per site).

As detailed in Appendix D below, the INL reviewers think the chi-square test should not be performed if there are fewer than three failures. This guidance is consistent with guidelines in Ref. [8]. A difference among sites can only be seen if multiple failures occurred at the same site and that site had relatively few demands or run hours. An example of evaluating a situation with very sparse data is included in Appendix D.

If a statistical test shows that differences can be seen, the hypothesis of poolability is rejected, and fitting an EB distribution is attempted. Otherwise, the uncertainty distribution comes from the JN or WIP methods.

The language describing the findings in the PWROG report was imprecise. See Appendix D in this report for further details.

4.2 Outlier Evaluation Summary

Appendix E below discusses the search for possible outliers, as discussed in Appendix B of the PWROG report. A nonparametric "Box Plot" method was used to see if data from one or more sites were atypical of the other data. That method found four instances in which data for certain sites exceeded the "upper fence" defined in the PWROG report. In each case, the outlier assessment was discounted.

Review of the PWROG report's Appendix B also raises concerns that far too many data were derived from only a few components. Data for these components were collected during 2008–2014, before the establishment of the current FLEX program. Such data biases the dataset if anything is atypical about those components. Concern also exists over the quality of testing and completeness of the old data.

4.3 Clarity and Level of Detail in Report

The need for more detail in certain areas has been mentioned. Particularly, the alpha values and specific references for each installed component distribution (used as a starting point for the WIP) should be included. The multiplier and E_M factor should also be specified. In the middle of p. 5-10, the report refers to "pump and non-pump components" instead of stating clearly that the multiplier ("4") applies to all 32 FLEX component uncertainty distributions.

The rest of the paragraph following the equations is also vague. An "initial distribution" is referenced. The term "nominal prior" is also used. They likely mean the same thing (i.e., the distribution from the 2015 Update). Then, factors are applied to generate a "scaled" distribution, but that distribution (if it includes the E_M variance factor) is the actual WIP distribution (the weakly informed prior). Defining these terms or using more specific terms would help readers more clearly understand the ideas being presented.

The RF is discussed on p. 5-11. The text suggests that it lie between five and ten. Further justification for this choice is desirable. The reviewers presume the reason relates to EFs used in other risk assessments. See the discussion in Appendix B for more details.

In the paragraph at the bottom of p. 5-11, use of the RF ratios of the final posterior distributions is discussed. Apparently, they should lie between 5 and 10 also, rather than being smaller. But data with many demands or operating hours along with several failures can easily lead to narrow posterior distributions. If the data are valid and representative of what would happen in an actual challenge in which FLEX components are brought into use, the conclusion (including small RF values) stands. Finally, the narrow distributions might be caused by issues relating to data quality and the possible influence of excess reliance on old records. The paragraph might make clear whether the "some cases" mentioned were identified.

The issue of homogeneity of the observed data is discussed on p. 5-11 in the paragraph about posterior distributions. Homogeneity is indeed required, because it is assumed by the Bayesian update. The reviewers do not think the property of being homogeneous is what makes the posterior distribution converge.

The report needs to be consistent about the rule for using the JN approach. On p. 5-6, the guidance says to use JN if the test for homogeneity is not rejected and the EB method fails. At the end of p. 5-7 and at the top of p. 5-8, the criterion for JN use depends on the amount of operating time or demands.

The reviewers question the idea expressed on p. 5-7 about the JN distribution being "too strong as a starting point." The JN prior is wide, but, as noted above, the posterior distribution can indeed be narrow after being updated with data containing several failures. The posterior distribution does converge when there is enough observed data. Perhaps the issue lies in the fact that the JN prior starts the analysis with the equivalent of 0.5 failures in 0 hours (for FTR) or in 1 demand (for FTS). "Too pessimistic" might be a more accurate description than "too strong."

Clarity is also discussed in Appendix D, in reference to the PWROG appendix on pooling.

5. SUMMARY AND CONCLUSIONS

The PWROG report documents a very credible effort and a substantial process to obtain uncertainty distributions for FLEX components, which are portable and brought wherever needed. WIP distributions were used for those FLEX components with relatively little data. The distributions were made "weakly informative" by (1) making the mean values four times higher than the mean values associated with similar installed equipment, and (2) multiplying the variance by a factor that widened the resulting prior distribution.

INL's evaluation of this process considered the data itself, the choices for installed equipment, the use of the factor inflating the mean values of the distributions, and use of the variance inflation factor. Suggestions for improvements were found in the following areas:

- In the data, records prior to a plant's implementation of the FLEX program should be excluded unless testing was as rigorous as current testing; event dates are needed for each failure; and the testing for instances cited as successes must be certified as similar to an actual demand.
- Consideration should be given to reducing the number of component type categories. Especially when data are sparse in certain categories, perhaps data across operating conditions should be pooled. In some cases, pooling across component type (for example, motor-driven pumps vs. positive displacement pumps) might also be beneficial.
- In a few cases, the 2015 Update data for installed components could be matched more closely to the portable components (such as based on component type; for example, using positive displacement pump data instead of centrifugal pump data for positive displacement pumps).
- For generator failure to run and pump failure to run, the 2015 Update dataset for installed components contains estimates for two running periods (for DG, FTLR and FTR; for pumps, FTR<1H and FTR>1H). The installed data would better match the FLEX data if the earlier period data were almost always used. The later period rate could be combined for tests lasting over an hour (i.e., for a two-hour test, the average of the two rates could be used).
- Evidence from components with a lot of data could be used to confirm or refute the factor of 4 for adjusting the mean, which at present is based entirely on engineering judgment. However, this assumes that the components with a lot of data are comparable to components with less data, not deserving exclusion because of age and resulting differences in failure probabilities or rates. Balancing the benefit of quantity with the danger of irrelevance is beyond the scope of this review.
- The method for choosing E_M should be changed so that the prior distribution is not chosen based on the posterior results. The method also needs to be changed due to cases in which no value of E_M gives the desired results. E_M should also be determined via a documented, objective process. An acceptable alternative would be to eliminate E_M and WIP altogether, and use the constrained noninformative prior matched to some multiple of the source data mean.
- The goal should be a realistic or conservative prior distribution. The posterior distribution, then, is whatever may result when the prior is updated by the FLEX data. There should be no attempt to directly adjust this posterior distribution.
- If the WIP distribution is retained, the alpha parameters for the input source distributions need to be specified in addition to the mean values. The report should also present the E_M factor for each WIP distribution. The process for identifying factors to inflate the distributions and make them less informative could be explained more clearly.
- Corrections in the PWROG Appendix B (poolability) would improve the report.

The study of the output WIP distributions and possible alternatives showed that the cutoffs for using WIP are reasonable. Of course, other distributions are possible. Compared with the posterior distributions from the constrained noninformative and Jeffreys prior distributions presented here, the posterior means from the WIP distributions were closer to the adjusted "4 times the nominal" values. They were also somewhat narrower for FTR.

6. REFERENCES

[1] Degonish, M.M., 2020, *FLEX Equipment Data Collection and Analysis*, PWROG-1803-P, Rev. 0.

[2] Haessler, R., 2016, *Implementation of FLEX Equipment in Plant-Specific PRA Models,* PWROG-14003-NP, Rev. 0.

[3] Eide, S, et al., 2007, *Industry-Average Performance for Components and Initiating Events at U.S. Commercial Nuclear Power Plants,* NUREG/CR-6928.

[4] 2015 Update: *Parameter Estimates Spreadsheet 2015* downloaded from https://nrcoe.inl.gov/resultsdb/AvgPerf/, accessed 4/8/2020.

[5] Teleconference with staff of NRC, PWROG, and INL, April 21, 2020.

[6] Tversky, A, & Kahneman, D. 1974, "Judgment under uncertainty: Heuristics and biases," in *Science*, Vol. 185, pp. 1124-1131.

[7] Block, R.A. & Harper, D.R., 1991, "Overconfidence in Estimation: Testing the Anchoring-and-Adjustment Hypothesis," in *Organizational Behavior and Human Decision Processes*, Vol. 49, pp. 188-207.

[8] Atwood, C.L., LaChance, J.L., Martz, H.F., et al., 2003, *Handbook of Parameter Estimation for Probabilistic Risk Assessment,* NUREG/CR-6823.

APPENDIX A: WIP AND CONSTRAINED NONINFORMATIVE (CN) PRIOR

A-1. SUMMARY

The WIP distribution is somewhat more informative than the CN prior. As a result, compared to the CN prior, the WIP distribution produces somewhat narrower posterior distributions whose means are closer to the adjusted means of currently installed equipment. Numerical comparisons are given in the body of this paper.

The main points derived here from the mathematical portion are as follows:

- Although the gamma(α, β) has two parameters, it is α that governs the distribution's shape and the concentration of the distribution around the mean. The value $\alpha = 0.5$ is traditionally used for noninformative priors. If α is larger, the distribution is more concentrated around the mean; but if α is smaller, the lower percentiles of the distribution can become wildly unrealistic.
- Detailed properties of the beta distribution are best understood in the common case when the beta distribution is approximated by a gamma distribution.
- The WIP distribution is very similar to the CN distribution but seems to usually have a somewhat larger α. As a result, it is slightly less diffuse.

A-2. BACKGROUND REVIEW

WIP (see Ref. [A-1]) and the CN prior (see Ref. [A-2]) are both intended to be diffuse but to have a specified mean. Depending on the failure mode, they are either gamma or beta distributions. These distributions and the interpretations of their parameters are summarized here.

The simplest formulation is for a gamma(α, β) distribution when estimating the rate of FTR. However, a beta(α, β) distribution is used when estimating the probability of FTS. The two are closely related, because a beta(α, β) approaches a gamma(α, β) distribution when the mean is small and β is large. For reliable equipment with small mean failure probability and a moderate amount of data, the beta distributions are approximately equal to gamma distributions. Therefore, the gamma formulation will be emphasized in the discussion below, with additional discussion of beta distributions as necessary.

The mean and standard deviation of a gamma(α, β) distribution are

$$\mu = \alpha/\beta \qquad \text{(A-1)}$$

and

$$\sigma^2 = \alpha/\beta^2 \qquad \text{(A-2)}$$

If λ has a gamma(α, β) distribution, the equation for the distribution can be manipulated to show that λ/β has a gamma(α, 1) distribution. That is, β changes the scale of the distribution but not the shape. In particular, changing β causes the mean to change, but not the range factor, defined in Ref. [A-1] as

$$\text{RF} = \text{sqrt}[\ \text{gamma}_{0.95}(\alpha, \beta) \ /\ \text{gamma}_{0.05}(\alpha, \beta)\ ]$$

Here, the subscripts identify the 95th and 5th percentiles of the distribution. To see that β does not affect RF, write RF as

= sqrt[[$\text{gamma}_{0.95}(\alpha,1)/\beta$] / [$\text{gamma}_{0.05}(\alpha,1)/\beta$]

The βs cancel each other, and the quotient depends only on α.

Finally, one way to eliminate the effect of scaling is to measure everything relative to the mean. The quantity σ/μ is called the "coefficient of variation." From the above formulas for μ and σ, it can be seen that the coefficient of variation is $1/\alpha^{1/2}$. Thus, when the mean is fixed, a small α corresponds to a large standard deviation; and as α approaches 0 while the mean is fixed, the standard deviation increases without bound.

Figure A-1, from Ref. [A-3], shows the gamma distribution corresponding to four values of α. Observe that when $\alpha < 1$, the density is unbounded near 0; and when $\alpha > 1$, the density has a skewed bell shape. A small α corresponds to a diffuse distribution (large standard deviation), while a large α corresponds to a distribution concentrated around the mean.

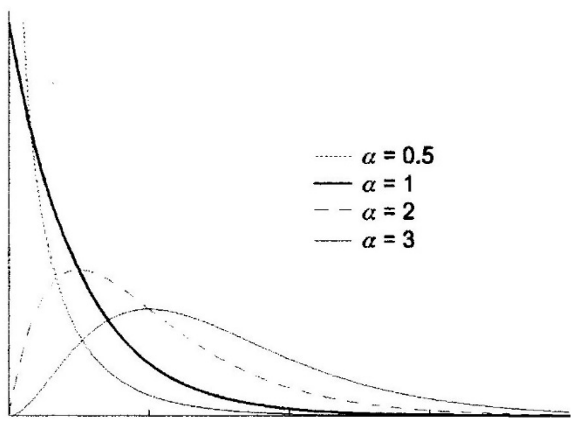

Figure A- 1. Four gamma densities with different shapes

For a beta distribution, the above formulas must be slightly modified. For this distribution we have

$$\mu = \frac{\alpha}{\alpha + \beta}$$

and

$$\sigma^2 = \frac{\mu(1-\mu)}{(\alpha+\beta+1)}$$

These are not quite the same as for the gamma distribution; however if μ is small (~1E-2) and α is much larger (> 0.4), the two equations approximately agree. Regardless of whether they agree, the equations show that when μ is held constant and α increases, σ decreases. This is the same relation in both cases.

For a gamma distribution, it is stated above that β is a scale parameter with no influence on RF. For a beta distribution, this is not strictly true. But if the beta distribution can be well-approximated by a gamma distribution, the statements about the gamma distribution are approximately true for the beta.

A-3. APPLICATION TO WIP DATA

In the FTR case, the CN prior is defined as a gamma(α, β) distribution with $\alpha = 0.5$ and β chosen to give the distribution the desired mean. In the FTS case, the CN prior is a little more complicated: it is a beta(α, β) distribution with α slightly less than 0.5 and β chosen to give the desired mean. The exact formulas are given in Refs. [A-2] and [A-3]. In each case, the mean is determined from some external source; then the CN prior is constructed to have that mean.

The CN prior imitates the Jeffreys noninformative (JN) prior, which is designed to maximize entropy (a theoretical measure of diffuseness) when the distribution is properly parameterized. To do this, the JN prior sets $\alpha = 0.5$ and $\beta = 0$ (for a gamma distribution) or $\beta = 0.5$ (for a beta distribution). As with the JN prior, the CN prior is designed to maximize entropy, but it is subject to the constraint on the mean. This is why both distributions use α at or near 0.5. It is good to see that the WIP, developed through engineering considerations, is not too different from the CN prior, which was developed through mathematical considerations.

If α is small, care must be taken when interpreting the lower percentiles of either distribution. This situation can arise if the prior α is small and no failures are observed in the data, making the posterior α also small. If α is much less than 1.0, the 5^{th} percentile may be unrealistically small by orders of magnitude. However, this value is just an artifact of the model built with a gamma or beta distribution. When no failure has been observed, the data offer no way to know how small the true rate or probability might be. Unrealistic percentiles problem would not arise if a different prior with larger α were used, or if a differently shaped prior distribution such as a lognormal were used. See Sec. 8.4.3 of Ref. [A-3] for a fuller discussion.

A-4. COMPARISON OF WIP AND CN RESULTS FOR FLEX DATA

In the body of this report, the WIP and two CN distributions are compared for the eleven cases presented in Ref. [A-1], seven FTR cases with gamma distributions, and four FTS cases with beta distributions. For each comparison, the assumed mean for one of the CN distributions was made equal to the WIP mean. By definition of the gamma CN distribution, the prior α always equals 0.5 (for a gamma distribution) or is slightly less than 0.5 (for a beta distribution with small mean).

Both the WIP and CN distributions use 4 times the industry mean as their mean. However, they differ in the treatment of the variance. CN has its variance built-in to maximize the entropy, subject to the constraint on the mean. On the other hand, for a variance, the WIP distributions use a multiplier of the industry component source variance. As a result, both parameters of each WIP distribution are free to vary, as determined by the desired mean and variance.

The statements below apply to the gamma distribution. They also apply to the beta distribution with small mean, because that beta distribution is approximately a gamma.

Consider a set of FLEX FTR data, and two possible prior distributions, one WIP and one CN. Let us compare these two prior distributions. They are based on the same adjusted data for installed components, so they are constructed to have the same mean, denoted μ. Consider the common case in which the WIP distribution has the larger α. By Eq. (A-1), the WIP distribution must also have a larger β. By Eq. (A-2), the variance is μ/β, so the WIP distribution also has the smaller variance.

Now consider the update formulas. It can be shown that when f failures in time t are used to update a gamma(α, β) prior, the posterior distribution is gamma($\alpha + f$, $\beta + t$). Therefore, the posterior mean is ($\alpha + f$)/($\beta + t$). This lies between the prior mean and the mean of the data. In fact, direct algebra shows that

$$\frac{\alpha + f}{\beta + t} = U\frac{\alpha}{\beta} + (1 - U)\frac{f}{t}$$

where

$$U = \frac{\beta}{\beta + t} \ .$$

The size of U governs how strongly the posterior mean is pulled toward the prior mean. Continuing the example in which the WIP distribution has the larger β, the WIP will also have the larger U, and therefore it pulls the posterior mean closer to the prior mean than the CN distribution does.

The relative strength of the two priors ultimately depends on α. This is seen in each of the nine WIP cases in this study that have FLEX failure data, even the FTS cases. If the WIP has a larger α than the CN prior, the estimate follows the FLEX data more closely when the CN prior is used, and it adheres closer to the adjusted industry mean when the WIP is used.

References

[A-1] Degonish, M.M., 2020, *FLEX Equipment Data Collection and Analysis*, PWROG-1803-P, Rev. 0.

[A-2] Atwood, C.L., 1996, "Constrained Noninformative Priors in Risk Assessment," in *Reliability Engineering and System Safety*, Vol 53, Issue 1, pp. 37-46.

[A-3] Atwood, C.L., LaChance, J.L., Martz, H.F., et al., 2003, *Handbook of Parameter Estimation for Probabilistic Risk Assessment,* NUREG/CR-6823.

APPENDIX B: ERROR FACTORS AND RANGE FACTORS

B-1. ALTERNATIVE DEFINITIONS OF ERROR FACTORS

The discussion in this appendix applies to gamma and lognormal distributions. It does not apply to beta distributions except in cases where the beta is closely approximated by a gamma.

The range factor (RF) for a gamma distribution seems patterned in the report after the error factor (EF), which is used with the lognormal distribution. Two definitions of EF are worth considering, both based on ratios of percentiles:

EF1 = sqrt($95^{th}/5^{th}$) and EF2 = $95^{th}/50^{th}$

For a lognormal distribution, the two expressions are numerically identical, since the logarithm of a lognormal random variable is normal and, therefore, symmetrical about the median. For a gamma distribution, these could be replaced by the RF, but the corresponding two terms are definitely not equal:

RF1 = sqrt($95^{th}/5^{th}$) and RF2 = $95^{th}/50^{th}$

As discussed in Appendix A, a ratio of percentiles of a gamma(α, β) distribution depends only on α, not β. Also, when α is small ($\ll$ 1.0), the lower percentiles can be orders of magnitude less than the mean. This is not an important issue for risk assessment, because it is the upper percentiles that are of concern. However, RF1 may be artificially inflated because it divides by the 5^{th} percentile, which is potentially very small. On the other hand, RF2 involves only percentiles of interest, the median and the 95^{th} percentile. Table B-1 shows some possible values of α and the corresponding two definitions of RF. For realistic comparisons, each row is constructed with the same mean, set to 1.0 for simplicity's sake.

Table B- 1 Two definitions of RF for gamma distributions with selected values of alpha and mean = 1.0

	Percentiles			Possible RF Definitions	
alpha	5^{th}	50^{th}	95^{th}	RF1 = sqrt($95^{th}/5^{th}$)	RF2 = $95^{th}/50^{th}$
0.84	0.0318	0.6421	3.1876	10.0099	4.9644
0.5	0.0039	0.4549	3.8415	31.2560	8.4439
0.4395	0.0019	0.4020	4.0208	46.0830	10.0008

As can be seen, if we use RF1, the definition involving the 5^{th} percentile, an RF of 10 corresponds to α = 0.84. However, if we use RF2, involving only percentiles of concern, an RF of 10 requires a smaller α—as small as approximately 0.44. A compromise would correspond to a CN prior, with α = 0.5. See Appendix A for further discussion of this topic.

B-2. INTERPRETATION AND USE OF ERROR FACTORS

Reference [B-1] gives the table on which Table B-2 (below) is based. It is intended for lognormal uncertainties of point estimates, for space launch vehicle applications. It is not authoritative, but provides an interesting example of how other, non-nuclear engineering applications use EFs.

Table B- 2 . Example of error factors in the aerospace industry

Source	Source Description	Source Application	Error Factor
Legacy Hardware	Other Launch Vehicle Data (Most Applicable)	Same component	3
		Like component	4
	Aerospace Data	Same component	5
		Like component	6
	Other Industry Data	Same component	6
		Like component	7
New Hardware	MIL-HDBK-217F Methods	Same component	8
		Like component	9
	Non-Expert Engineering Judgment (Least Applicable)	Documented Process	10
		Undocumented Process	15

Reference

[B-1] Hassan, M.A., Novack, S., and Ring, B., 2016, "Source Data Applicability Impacts on Epistemic Uncertainty for Launch Vehicle Fault Tree Models," downloaded from https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20160007066.pdf.

APPENDIX C: EVALUATION OF DEMAND/OPERATING HOUR CUTOFFS

C-1. SUMMARY

To facilitate a study of the cutoff values separating the modeling of FLEX component uncertainties using updated WIPs from the modeling using updated JN distributions, probabilities and rates from both modeling methods were plotted together. FTR rates were plotted as a function of run hours, while FTS probabilities were plotted as a function of demands. In each plot, reference lines (at demands = 50 for FTS, and hours = 100 for FTR) separate the WIP data from the JN data. For points on the left side of each plot, the PWROG study reported WIP distributions; on the right, it reported JN (and EB) distributions for components with greater operating experience. For the purpose of this study, the EB component distributions were not relevant.

For both failure modes, increasing the operating data cutoff for WIP would allow more component distributions to be treated with the WIP method rather than the JN method, and vice versa. Therefore, the WIP method was applied to the components with JN distributions, and the JN method was applied to the components with little operating experience. The plots show the mean values and 95th percentiles for both distribution methods.

An error was noted in the fitting of the adjusted WIP distributions to beta distributions (for FTS). The INL reviewers re-computed the means and percentiles for the four FTS WIP distributions in the study.

The plots show that the JN means and 95th percentiles are slightly larger than the WIP values for all but one component not modeled via the EB method. This one exception is for FTR with a component that had 119 run hours—just over the 100-hour cutoff. Its uncertainty distribution would have a higher mean and 95th percentile with the WIP distribution, but all others are higher with the JN distribution. The differences shrink on the right side of the plots as operating experience increases. Generally, decreasing the cutoffs and having more JN distributions slightly increases the mean values and 95th percentiles.

C-2. BACKGROUND

Viewing the FLEX distribution means and 95th percentiles sorted by number of demands or run time provides an overview of the results of the WIP and JN methods used within the PWROG report to establish uncertainty distributions. In the report, the EB method was used if sufficient data existed for the maximum likelihood algorithm to converge (and if statistical tests reject the hypothesis that the data are homogeneous across sites). When the EB method was unsuitable, the PWROG analysts used data updates of the JN distribution or WIP to estimate distributions. The cutoff values separate the WIP component distributions from the JN ones. The EB distributions are not a part of this study.

The WIP data were for those components with the least operational data. External data were needed to develop distributions for these data. Distributions for components with more data were processed using the JN method. The cutoffs between these two methods were initially set at 50 demands and 100 run hours, apparently based on engineering judgement. The purpose of this appendix is to observe the results of these decisions and evaluate the WIP/JN cutoff points.

C-3. ANALYSIS METHOD

To make the comparisons, WIP distributions were needed for the situations treated with JN distributions, and JN distributions were needed for the components with few demands or operating hours.

Since the JN distributions depend only on the data, they are easy to reconstruct for those components with little data. For FTS, the noninformative beta (0.5, 0.5) distribution is updated with the data; for FTR, the noninformative gamma (0.5, 0) distribution is updated. The formulas on p. 5-7 in the PWROG report were used.

For the WIP distributions, the process is more involved. The PWROG study identified data about installed components for each type of FLEX component. These are the “nominal” distributions, with specified values for the mean, alpha parameter, and beta parameter. (Note that the α and β for FTS are for beta distributions, while the α and β for FTR are for gamma distributions.)

The next step is to make the distributions “weakly informed.” The PWROG team multiplied all the mean values and standard deviations by 4 to obtain a “scaled” distribution. To provide additional flexibility to widen the WIP they introduced another multiplier to increase the variance. Initially, the variance multiplier was 1.0.

The gamma or beta distributions that had the desired mean and variance were identified. These WIP or scaled distributions were updated with data to generate posterior distributions for the portable equipment having little data.

As a final step, the WIP posterior distributions were reviewed to ensure that they were “wide enough”. The RFs, defined as the square root of the ratio of the 95^{th} percentile to the 5^{th}, were computed. The goal was to have the RF of the *posterior* distribution resulting from a WIP to be between 5 and 10. Often, increasing the variance multiplier was required in order to have a higher posterior RF. By trial-and-error, as needed, each variance multiplier was increased to widen the corresponding WIP distributions. The final reported distributions for use in risk assessments were the process’s posterior distribution outputs with the adjusted variance multipliers.

INL processed these same steps for the components with more operating experience that the PWROG modeled with JN distributions.

C-4. THE FTS ERROR

In the process of calculating a WIP distribution for each JN component and a JN distribution for each WIP component, the WIP and JN calculations in the report were checked. Errors were found in the four WIP calculations for FTS (the beta distribution calculations). The errors are shown in Table C-1. The stated means and variances agree in both the INL and PWROG calculations. However, the α and β values do not agree, and the variance of the distributions having the PWROG (α, β) values do not have the stated variance. It appears that the PWROG α values are similar to the INL β values. In any case, the calculations result in both the α and β values decreasing by more than a factor of 10. The results depend heavily on the choice of the variance factor (E_M) that PWROG personnel selected. The selected values are shown in the table.

When the α or β parameter of a beta distribution is so close to zero, the behavior of the distribution changes. Note that the first two components in Table C-1 have the same data, based on the same nominal (source) information, and no failures. Therefore, the first component is ignored in the following observation. The 5^{th} percentiles for the three distinct WIP distributions go from numbers around 1E-3

(9.0E-4, 9.7E-4, and 1.12E-3, respectively) to 4.41E-107, 1.72E-90, and 3.46E-69, respectively. As a result, the WIP (prior distribution) RFs defined by the PWROG go from values of 5–7 to, respectively, 1.5E+52, 1.1E+44, and 4E+33. No adjusting of the variance factor fixes this situation.

Table C-1. PWROG and INL values for WIP distributions for FTS.[a]

Calculations	Portable Component Acronym	Scaled Mean	Scaled Variance	Adjusted or Scaled α	Adjusted or Scaled β	E_M
For diesel-driven positive displacement pumps with low pressure, high flow (no failures)						
PWROG	PDDPDP-LPHF	8.68E-03	3.58E-03	1.39	159.0	30
INL	PDP_D_LP	8.68E-03	3.57E-03	0.0123	1.400	30
For diesel-driven positive displacement pumps with medium pressure, med. flow (no failures)						
PWROG	PDDPDP-MPMF	8.68E-03	3.58E-03	1.39	159.0	30
INL	PDP_D_MP	8.68E-03	3.57E-03	0.0123	1.400	30
For diesel generators with low voltage (2 failures)						
PWROG	PDG-LV	1.15E-02	5.04E-03	1.24	107.0	350
INL	GEN_D_LV	1.15E-02	5.04E-03	0.0145	1.246	350
For motor-driven air compressors (1 failure)						
PWROG	PMDAC	1.66E-02	7.67E-03	1.12	65.9	35
INL	CPR_M	1.66E-02	7.63E-03	0.0190	1.124	35
a. All distributions in this table are prior distributions. All the E_M values are from the PWROG.						

The PWROG method uses the E_M factor to try to make the RF of the final *posterior* distribution be between 5 and 10. For the first two components in Table C-1, with no failures, the best that can be done is to set $E_M = 1$. The posterior distribution RF is around 18 and will go no lower. The posterior distribution α parameter is not near zero for the other two components that have failures. However, setting $E_M = 35$ for the portable diesels, along with 2 failures, makes the associated RF 3.5, and it will go no higher. For the air compressors, an E_M value of 35 results in the posterior RF being 7.3, which satisfies the goal.

The conclusion is that the PWROG method fails to control the posterior distribution RF for the beta distributions a significant proportion of the time. This fact, together with the impropriety of adjusting the prior after looking at the data (as reflected in the posterior distribution), makes the method untenable.

However, for the purpose of assessing the somewhat moot effect of the cutoff values for choosing between WIP and JN methods in the PWROG study, the INL reviewers applied the above procedure (with correct formulas for obtaining the WIP beta distributions) to all the non-EB FTS distributions. The procedure was also applied to all the FTR distributions treated with the JN method.

C-5. RESULTS

The mean and α parameters for all WIP and JN posterior distributions are presented in Tables C-2 through C-5. Tables C-2 and C-3 provide posterior WIP estimates for FTS and FTR, respectively, and the other two tables do the same for JN data. Each table has a horizontal line separating the components that PWROG analyzed with the WIP method from those with the JN method. The data in the rows in Tables C-3 with PWROG Method = “WIP” and the data in Tables C-4 and C-5 with PWROG Method = “JN” are the same as the PWROG results. The remaining tabulated information, including all of Table C-2, was calculated by the INL reviewers only.

Table C-2. WIP posterior beta distribution parameters for probability of FTS[a].

FLEX Component	Dem.	PWROG Method	WIP Posterior Beta Distributions for Uncertainty in FTS						
			Mean	α	5th	50th	95th	Variance Multiplier (E_M)[b]	Range Factor
PDP_D_LP	0	WIP	8.68E-03	0.619	9.45E-05	4.71E-03	3.07E-02	1	18.04
PDP_D_MP	22		6.63E-03	0.619	7.21E-05	3.59E-03	2.35E-02	35	18.1
GEN_D_LV	27		4.63E-02	2.249	9.74E-03	4.02E-02	1.04E-01	3	3.3
CPR_M	34		2.41E-02	1.399	2.56E-03	1.89E-02	6.36E-02	2	5.0
PDP_M_LP	57	JN	2.59E-03	1.100	1.67E-04	1.87E-03	7.50E-03	45	6.7
PDP_M_MP	78		2.19E-02	2.046	4.06E-03	1.86E-02	5.11E-02	1	3.5
PDP_D_HP	133		3.03E-03	0.619	3.28E-05	1.64E-03	1.08E-02	2	18.1
GEN_T_HV	146		8.48E-03	1.248	7.18E-04	6.38E-03	2.34E-02	2	5.7
GEN_T_MV	158		3.92E-02	6.248	1.77E-02	3.73E-02	6.73E-02	25	2.0
GEN_D_HV	176		6.55E-03	1.353	6.45E-04	5.04E-03	1.76E-02	45	5.2
CPR_D	293		2.09E-02	6.346	9.44E-03	1.99E-02	3.59E-02	65	2.0
PDP_M_HP	507		5.86E-03	3.031	1.62E-03	5.24E-03	1.22E-02	1	2.8

a. The data for the diesel-driven positive displacement pumps with low pressure/high flow rate (PDP_D_LP) are not plotted; there were no demands.
b. All the E_M values were estimated by INL using the PWROG method for WIP distributions.

Table C-3. WIP posterior gamma distribution parameters for rate of FTR[a].

FLEX Component	Run Hrs.	PWROG Method	WIP Posterior Gamma Distributions for Uncertainty in FTR						
			Mean	α	5th	50th	95th	Variance Multiplier (E_M)[b]	Range Factor
PDP_D_LP	0	WIP	3.92E-03	1.05	2.25E-04	2.77E-03	1.16E-02	10	7.2
GEN_D_LV	8.9		5.83E-03	1.26	4.98E-04	4.38E-03	1.61E-02	2	5.7
PDP_D_MP	17		3.69E-03	1.05	2.12E-04	2.60E-03	1.09E-02	10	7.2
CPR_M	19		5.36E-05	1.13	3.63E-06	3.88E-05	1.54E-04	20	6.5
PDP_M_LP	27		4.80E-04	0.84	1.54E-05	3.09E-04	1.53E-03	1	10.0
PDP_M_MP	35.4		1.05E-03	1.84	1.67E-04	8.64E-04	2.55E-03	1	3.9
PDP_D_HP	97.5		2.87E-03	1.05	1.65E-04	2.03E-03	8.46E-03	10	7.2
GEN_T_MV	119.3	JN	2.31E-02	6.50	1.05E-02	2.19E-02	3.98E-02	1	1.9
GEN_T_HV	124.7		1.01E-02	1.37	1.01E-03	7.76E-03	2.71E-02	15	5.2
GEN_D_HV	175		5.81E-03	1.36	5.74E-04	4.46E-03	1.56E-02	7	5.2
PDP_M_HP	254.9		7.71E-03	2.00	1.37E-03	6.47E-03	1.83E-02	350	3.7
CNP_D_HP	306.3		7.63E-03	3.70	2.46E-03	6.96E-03	1.51E-02	15	2.5

a. The data for the diesel-driven positive displacement pumps with low pressure/high flow rate (PDP_D_LP) are not plotted; there were no run hours.

b. Variance multipliers for rows with PWROG method=”WIP” were estimated by the PWROG. Variance multipliers in the remaining rows were estimated by the INL reviewers using the PWROG method for WIP distributions.

Table C-4. JN beta distribution parameters for probability of FTS[a].

FLEX Component	Dem.	PWROG Method	JN Beta Distributions for Uncertainty in FTS				
			Mean	α	5th	50th	95th
PDP_D_LP	0	WIP	0.5	0.5	3.93E-03	5.00E-01	9.94E-01
PDP_D_MP	22		2.17E-02	0.5	8.74E-05	1.02E-02	8.27E-02
GEN_D_LV	27		8.93E-02	2.5	2.25E-02	7.96E-02	1.90E-01
CPR_M	34		4.29E-02	1.5	5.25E-03	3.44E-02	1.09E-01
PDP_M_LP	57	JN	8.62E-03	0.5	3.42E-05	3.97E-03	3.30E-02
PDP_M_MP	78		3.16E-02	2.5	7.49E-03	2.78E-02	6.92E-02
PDP_D_HP	133		3.73E-03	0.5	1.47E-05	1.71E-03	1.43E-02
GEN_T_HV	146		1.02E-02	1.5	1.21E-03	8.08E-03	2.65E-02
GEN_T_MV	158		4.09E-02	6.5	1.93E-02	3.90E-02	6.95E-02
GEN_D_HV	176		8.47E-03	1.5	1.00E-03	6.71E-03	2.20E-02
CPR_D	293		2.21E-02	6.5	1.02E-02	2.10E-02	3.78E-02
PDP_M_HP	507		6.89E-03	3.5	2.15E-03	6.25E-03	1.38E-02

a. The data for the diesel-driven positive displacement pumps with low pressure/high flow rate (PDP_D_LP) are not plotted because there were no demands.

Table C-5. JN gamma distribution parameters for rate of FTR[a].

FLEX Component	Run Hrs.	PWROG Method	JN Gamma Distributions for Uncertainty in FTR				
			Mean	α	5th	50th	95th
PDP_D_LP	0	WIP	—	—	(no result)	—	—
GEN_D_LV	8.9		5.62E-02	0.5	2.21E-04	2.56E-02	2.16E-01
PDP_D_MP	17		2.94E-02	0.5	1.16E-04	1.34E-02	1.13E-01
CPR_M	19		2.63E-02	0.5	1.03E-04	1.20E-02	1.01E-01
PDP_M_LP	27		1.85E-02	0.5	7.28E-05	8.42E-03	7.11E-02
PDP_M_MP	35.4		4.24E-02	0.5	4.97E-03	3.34E-02	1.10E-01
PDP_D_HP	97.5		5.13E-03	1.5	2.02E-05	2.33E-03	1.97E-02
GEN_T_MV	119.3	JN	1.26E-02	0.5	1.47E-03	9.92E-03	3.28E-02
GEN_T_HV	124.7		1.20E-02	1.5	1.41E-03	9.49E-03	3.13E-02
GEN_D_HV	175		8.57E-03	1.5	1.01E-03	6.76E-03	2.23E-02
PDP_M_HP	254.9		9.81E-03	1.5	2.25E-03	8.54E-03	2.17E-02
CNP_D_HP	306.3		1.14E-02	2.5	3.54E-03	1.04E-02	2.30E-02

a. The data for the diesel-driven positive displacement pumps with low pressure/high flow rate (PDP_D_LP) are not plotted because there were no run hours.

Figures C-1 and C-2 show FTS posterior distribution information as a function of the number of demands, while Figures C-3 and C-4 show the same information for FTR as a function of run hours. The first figure in each pair contains WIP and JN means, while the second contains WIP and JN 95th percentiles. In each of the plots, the PWROG employed the WIP method only for situations to the left side of the reference line and the JN method only for those to the right side of the reference line. The JN curve on the left and the WIP curve on the right show the effects of moving the cutoff values.

The JN distribution uses only the FLEX data, whereas the WIP distributions start with data for installed components that have been used for years in the NPP industry. The mean values for the industry data are thus typically lower than those that come from the FLEX data (even though those mean values are increased by a factor of 4). The dip in Figures C-3 and C-4 in the PWROG (WIP) data are caused by a 1.35E-5 industry value for motor-driven air compressors. It is unsurprising that lowering the cutoff values increases the estimated mean and standard deviation. Sometimes, the difference is greater than an order of magnitude. This situation holds true for both FTS and FTR.

Figure C-1. Posterior WIP and JN mean values for the probability of a component's failure to start as a function of the number of demands.

Figure C-2. 95th percentiles for posterior WIP and JN uncertainty distributions for the probability of a component's failure to start as a function of the number of demands.

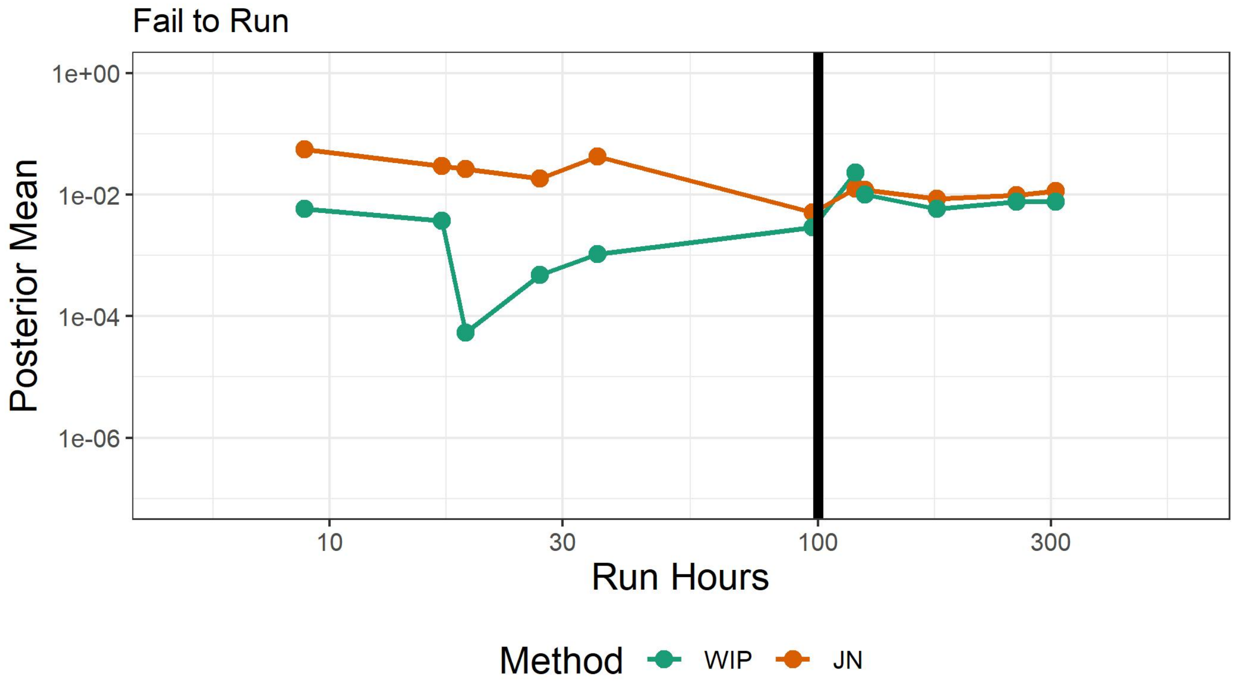

Figure C-3. Posterior WIP and JN mean values for the probability of a component's failure to run as a function of run hours.

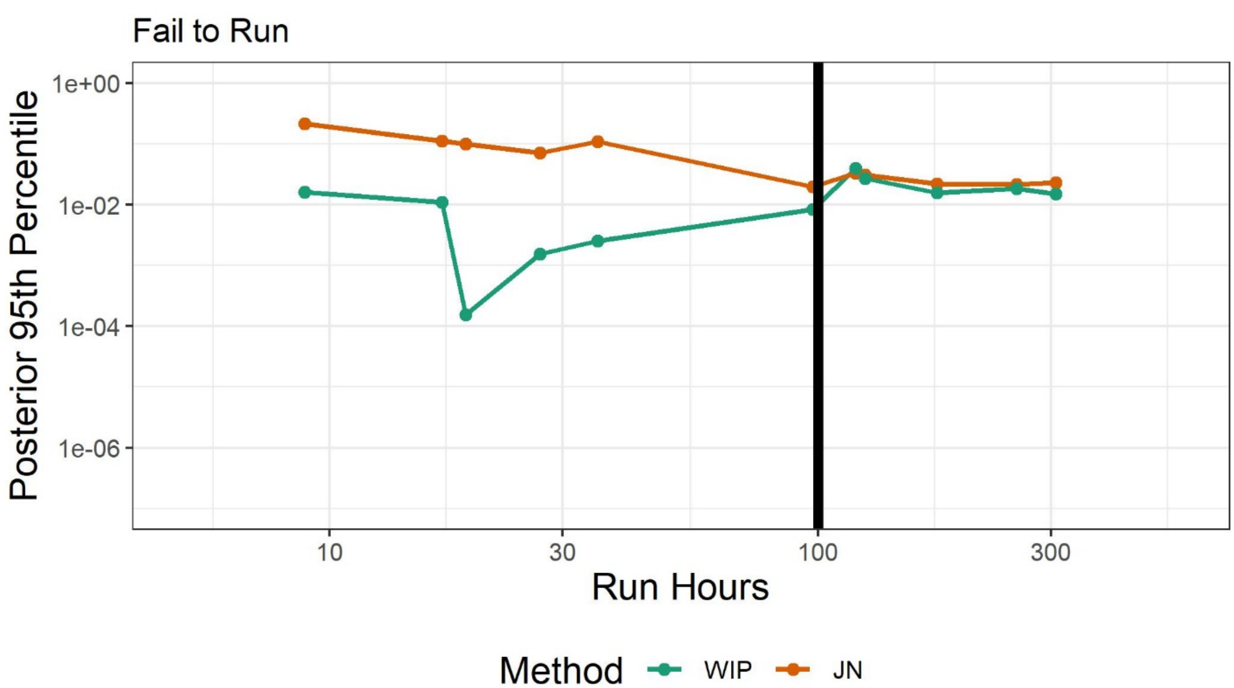

Figure C-4. 95^{th} *percentiles for posterior WIP and JN uncertainty distributions for the rate of a component's failure to run as a function of run hours.*

The 95th percentile plots show a similar pattern, with the JN values considerably higher than the WIP percentiles on the left. Even if the mean values were the same, the JN 95th percentile would exceed the WIP percentile any time the posterior JN α was less than the posterior WIP α parameter. Thus, decreasing the cutoff value for WIP increases the uncertainty estimates for the affected components.

In FTR Figures C-3 and C-4, within a small range above the cutoff points, the INL-estimated posterior WIP distribution values exceed the JN values. In that particular case (GEN_T_HV), increasing the cutoff so there will be more WIP distributions will bring in one higher outcome (the WIP mean is 2.31E-2, while the JN mean is 1.26E-2).

APPENDIX D: REVIEW OF POOLING STUDY

D-1. SUMMARY

The pooling study in Appendix A of Ref. [D-1] presents the data and conclusions for the statistical tests used to decide which statistical analysis should be performed for each dataset. However, unlike most of the PWROG report, this seems repetitive and largely unnecessary. In particular, many chi-squared tests are performed with datasets much too small for the tests to be valid. Nevertheless, conclusions are drawn, apparently using subjective, unstated criteria. These conclusions seem reasonable, but the argument to get there is muddled.

D-2. CHI-SQUARED TESTS WITH EXTREMELY SPARSE DATA

Ref. [D-1], p. 5-6, states, “The empirical Bayes analysis requires at least 3 failures in the dataset, otherwise the analysis ‘fails’”. So, what should be done with a dataset that has, at most, two failures—or indeed with any dataset when the EB method fails? There is no choice; it must be analyzed as if there is no heterogeneity among plants. In a way, the EB “failure” serves as a test, stating that there is insufficient evidence of between-plant difference to enable an EB model to be fit.

However, suppose a chi-squared test were attempted. The test statistic X^2 has an asymptotic chi-squared distribution if the dataset has a large number of both failures and successes. But what is the distribution if the dataset is small?

As an example, consider Section A.2.1.1 of Ref. [D-1], FTS of high-voltage diesel generators. (The data are reproduced in Table D-1). The hypothesis to be tested is

H_0: the probability of failure is the same on every demand.

As already mentioned, the large-sample approximate distribution of X^2 cannot be used. The only relevant distribution is that of X^2 *conditional on the fact that only one failure occurred*. Given that one failure occurred in 176 demands, there are exactly four possible outcomes, since that one failure could occur at any one of the four sites. Therefore, X^2 can only take four possible values. (This confirms that the distribution is not close to a chi-squared distribution.) The formulas for calculating X^2 in a 4 × 2 contingency table are given in many statistics texts, including Section 6.3.3.1.2 of Ref. [D-2]. A little work with a spreadsheet shows that the four possible values are 3.652, 8.828, 2.398, and 1.597, respectively, depending on which of the four sites had the failure.

Table D-1. Number of failures to start on demand, at four sites.

Site	Failures	Demands	Dems/Tot
S39	0	38	0.216
S52	1	18	0.102
S57	0	52	0.295
S58	0	68	0.386

Now, let us find the probabilities of these four possible values. Since only one failure occurred, H_0 says each demand has the same chance—namely, 1/176—of being the one failure. Therefore, the probability of the failure occurring at a particular site is proportional to the number of demands at that site. For example, the probability of failure occurring at S52 is 18/176 = 0.102. (This is the probability that $X^2 = 8.828$.) Similarly, the probabilities of the four possible values of X^2 are given in the right column of the Table D-1.

The failure actually occurred at S52, so the p-value for testing H_0 is 0.102. This is greater than 0.05, so the hypothesis of the same failure probability on every demand is not rejected. In fact, none of the outcomes have probability < 0.05, so there is no possible dataset with a single failure that would reject H_0 at significance level 0.05.

This conclusion can be compared with the result stated in Appendix A of the PWROG report. The calculated X^2 equals 8.828, which is greater than 7.815, the 95th percentile of the chi-squared distribution with 3 degrees of freedom. If only the calculated numbers were used, this would call for rejection of H_0. However, Appendix A of Ref. [D-1] says that the sample size is small, and since 8.23 is close to 7.815, H_0 should not be rejected. Instead, the data should be pooled. This is a very roundabout way of reaching a correct conclusion—one regrettably dependent on personal judgement.

We suggest saying instead that this is a small dataset—every cell has an expected failure count of < 0.5. Therefore, the data do not provide enough evidence to reject H_0 (i.e., poolability). Therefore, the data will be pooled because it is the simplest thing to do with such a sparse set of data. This conclusion applies to all the cases with fewer than three failures and fewer than 50 demands or 100 hours of running time. The borderline cases with just three failures may need special consideration.

D-3. MINOR ISSUES

D-3.1. Use of Two Significance Levels

Appendix A of Ref. [D-1] seems to use two significance levels: 0.05 and 0.10. It repeatedly talks about "a statistically significant difference…between sites at the 10% confidence level," but it judges this by whether or not the p-value from the chi-squared test is less than 0.05. For consistency, the reviewers recommend dropping all consideration of the 10% level. In most situations for testing whether a certain model is correct against a vague alternative, 5% is standard.

The report should state whatever rule is being used to decide if the dataset is poolable. However, it should skip the hypothesis-testing step altogether when the sample size is too small for the chi-squared test. There is no readily available small-sample test to use instead of the chi-squared test.

D-3.2 Use of Ambiguous, Nonstandard Statistical Language

The report *frequently* says "the test is accepted." But, using the metaphor of a legal trial, the test is the judge, and the judge's decision is whether or not to reject a certain claim. One doesn't accept or reject the judge, only the claim. Instead, say "poolability is accepted (or rejected)" or "H_0 is accepted (or rejected)." (Also, "accept" really means not to reject the default model due to insufficient evidence against it. As in a court of law, the defendant is guilty or not guilty, but "not guilty" does not necessarily mean "innocent"—just that there was insufficient evidence to convict.)

Moreover, the expression is used inconsistently. For example, in Sec. A.4.3.2 of Ref. [D-1], the conclusion is, "the p-value is significantly *less than* 0.05; therefore, the chi-squared *test is accepted*." (Italics ours.) But in Section A.7.1.1 of Ref. [D-1], the conclusion is, "the p-value is *larger than* 0.05, therefore the chi-squared *test is accepted*." One of these statements is evidently a misprint, but it is not clear which, since standard language is not used. Likewise, A.6.1 of Ref. [D-1] "accepts the test" because the p-value is small, and A.7.2.2 of Ref. [D-1] "accepts the test" because the p-value is large. To call this confusing is an understatement.

References

[D-1] Degonish, MM, 2020, *FLEX Equipment Data Collection and Analysis*, PWROG-1803-P, Rev. 0.

[D-2] Atwood, CL, LaChance, JL, Martz, HF, et al., 2003, *Handbook of Parameter Estimation for Probabilistic Risk Assessment,* NUREG/CR-6823.

APPENDIX E: REVIEW OF OUTLIER STUDY

E-1. SUMMARY OF OBSERVATIONS

The nonparametric "Box Plot" method was used to see if the rates or probabilities from one or more sites were atypical of the other data. That method found four instances in which data for certain sites exceeded the "upper fence" (UF) defined in the PWROG report. However, in each case, the outlier assessment was discounted because "the data were sparse."

Checking for outliers provides additional assurance that the development of uncertainty distributions for FLEX components was performed with due diligence. The decision to retain the points was conservative.

An evaluation of the data shows that some sites have much more operational experience than others. The validity of datasets from periods prior to the establishment of the FLEX program (approximately since 2012) is an issue. The tests must be thorough and well-documented, and all failures need to be reported. Even if the testing was adequate, the presence of huge amounts of data from only a few components will bias the resulting dataset if there are any abnormalities in those components.

If the old data were removed, the outlier investigation might have different results.

E-2. SUMMARY OF PWROG OUTLIERS APPENDIX

The PWROG study focused on whether there might be site-level outliers among the data for estimating FTR rates or FTS probabilities for FLEX components. For each estimate and site, the failure rate or probability was calculated as failures over time or failures over demands. The resulting ratios for each estimate were evaluated using the nonparametric "Box Plot" method. In each set of data, the median was identified (the center value, or the average of the two center values in an even number of sites). Then, the lower quartile was identified as the median of the data entries below the main median, and the upper quartile was the median of the data entries above the main median. Finally, the interquartile range was defined as the upper quartile minus the lower one. An "upper fence" (UF) was defined as the upper quartile plus 3 times the inter-quartile range. Values exceeding the UF were identified as possible outliers.

Potential outliers were identified for four of the 32 FLEX distributions. For each case, the data were listed along with the UF value and other aspects of the box plot. In each case, the potential outliers were judged to belong to the associated datasets despite being above the UF. The sparsity of the data was cited in each case as a reason to retain the flagged data.

E-3. REVIEW COMMENTS

This appendix provides graphs that display the associated data by plotting the number of failures as a function of the number of demands or operating hours (see Figures E-1 through E-4). In the graphs, the slope of a line from any point to the origin is the occurrence rate or probability corresponding to that point. The points with high values are on the left side of the graphs where the slope from a point to the origin is highest. In each figure, the diagonal line represents the UF. In each plot, points to the left of this line exceed the UF value.

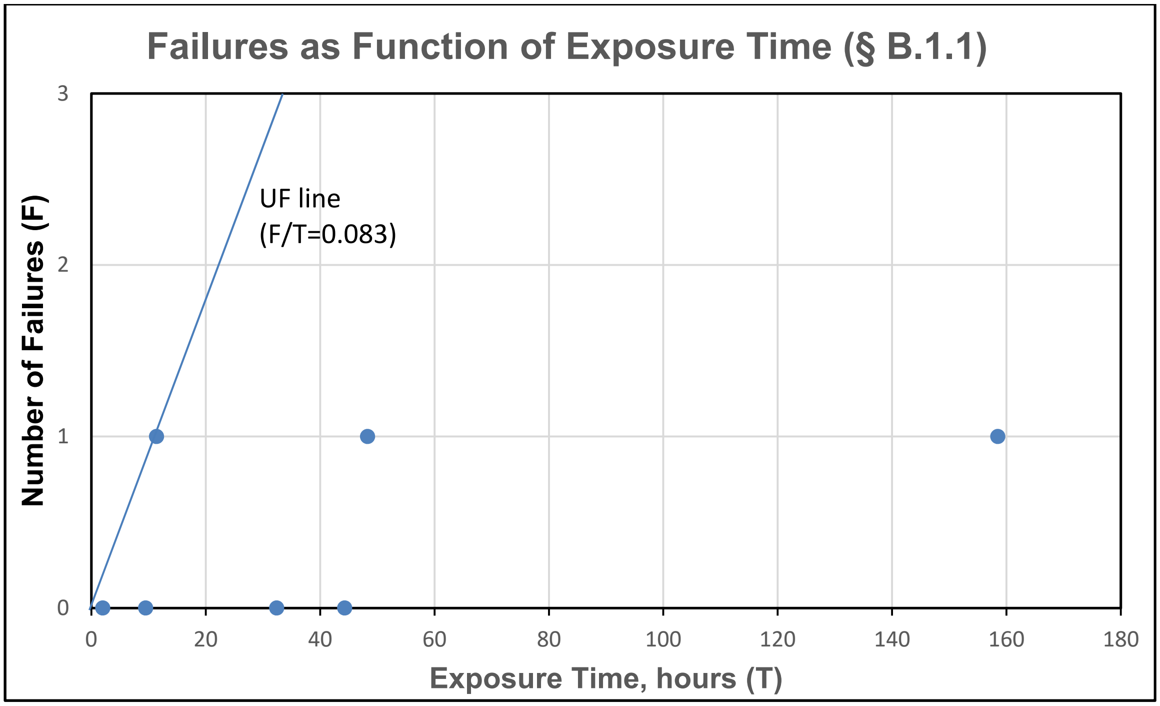

Figure E-1. Portable diesel-driven centrifugal pump (high pressure, low flow rate)—FTR.

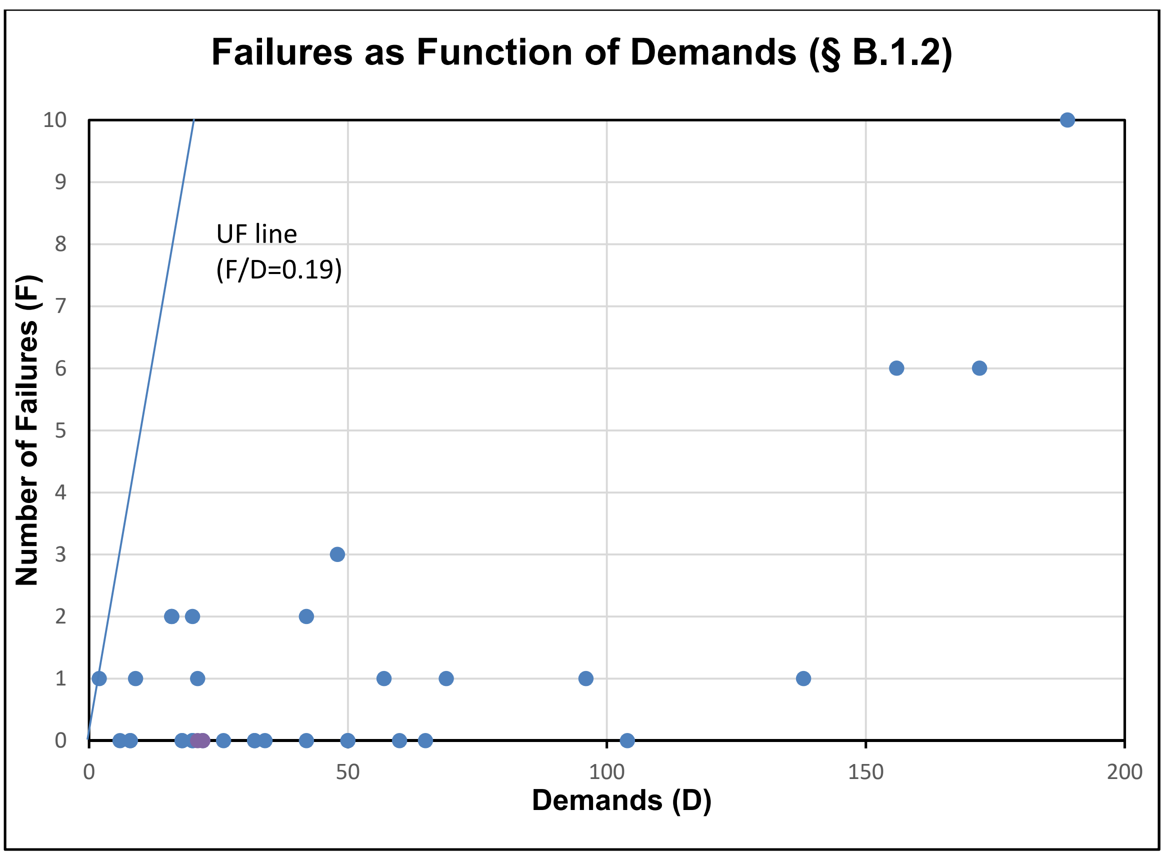

Figure E-2. Portable diesel-driven centrifugal pump (low pressure, high flow rate)—FTS.

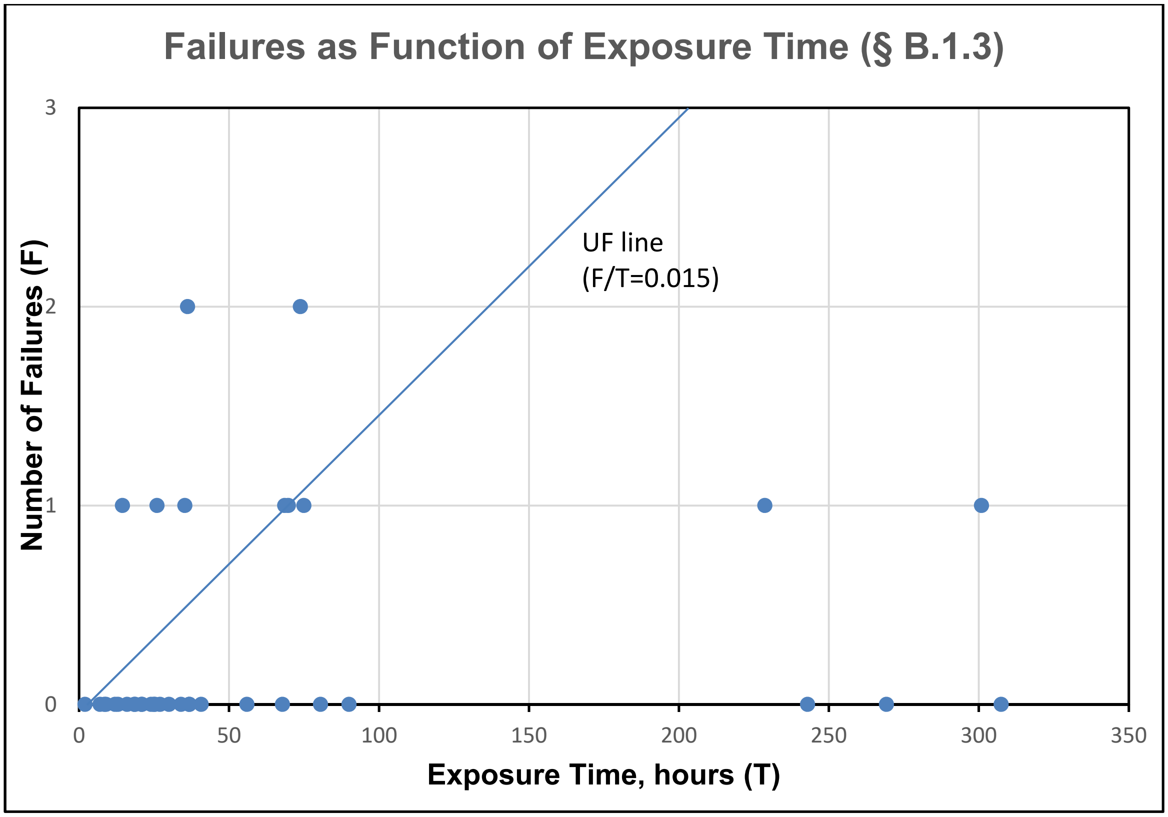

Figure E-3. Portable diesel-driven centrifugal pump (med. pressure, med. flow rate)—FTR.

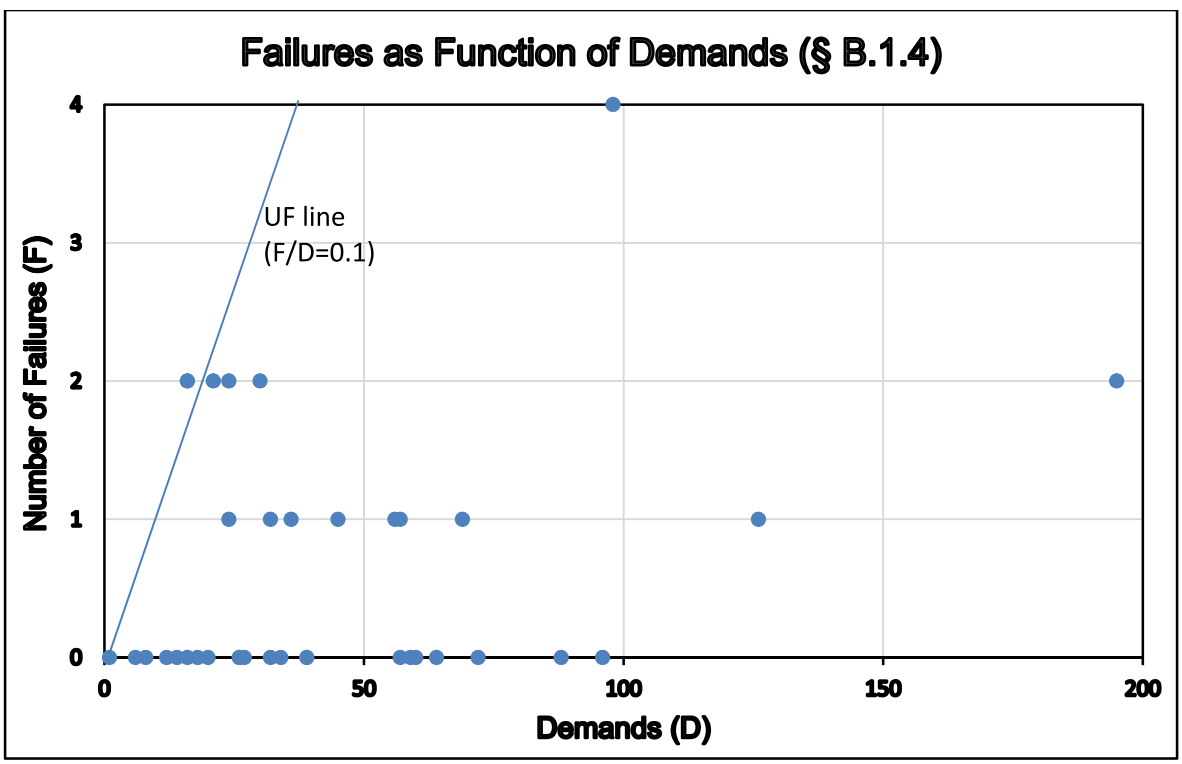

Figure E-4. Portable diesel-driven centrifugal pump (med. pressure, med. flow rate)—FTS.

In the first two plots, only one point exceeds the UF, and it has only one failure. The demands or exposure time are low. The third plot has five points to the left of the UF. The last has just one observed point but it is actually two sites, each with two failures and the same small number of demands.

Leaving these points in the datasets is useful because it is a conservative choice. From an informal point of view, they do not particularly stand out in the figures. If they were real outliers, it would behoove analysts to discover the reasons behind them. Further, unusual situations occur, and trying to rule out particular situations is not justified. The only real reason to rule out certain observations is if they contain errors such as misclassifications.

E-4. ADDITIONAL OBSERVATIONS FROM THE GRAPHS

Figures E-1 through E-4 show a different type of potential outlier. The aspect of the data that most stands out to the casual observer is the great disparity in exposure times and the numbers of demands from one facility to another. Some sites have data from as early as 2008, though guidance for implementing the FLEX program was not issued until 2012, and was modified as recently as December 2016. Inclusion of data from before the implementation of the FLEX program for these sites may be the primary cause of this disparity. Including these data makes the FLEX data unbalanced. The components at those sites have a much bigger influence on the outcome than do other components. If they are atypical in any way, the datasets will be biased. Excluding old data (which would apply only to the site itself, not the industry as a whole) could produce significant changes in the PWROG FLEX uncertainty distributions.

Another concern relates to the quality of testing at these sites. In more recent tests, a specification exists. It is not clear that tests prior to the adoption of FLEX standards were performed with the same rigor. Every test that contributed to the success count must challenge the components in the same manner as an operational situation. The reviewers suggest that the old data be excluded or else be validated.

The reviewers wonder if the same points identified as potential outliers in the PWROG report would have been flagged if data from the sites that had much larger amounts of data were filtered down to the same time period as the other sites.