NRC STAFF COMMENTS ON DRAFT NUREG-1829 AND RESPONSES

NRC STAFF COMMENTS ON DRAFT NUREG-1829 AND RESPONSES

General Comments

Comment Number: NGC1
Comment: There are many reasons to believe that the results of the expert elicitation process should not be used as a building block for modifying NRC's most important regulation affecting the design bases of nuclear power plants (i.e., 10CFR50.46). Although it is state-of-the-art, its uncertainties are so large (and understated) that its use in the regulatory arena should be deeply questioned.

Response: The commenter provides no basis for the assertion that the uncertainties are understated. Nevertheless, one of the initial analysis assumptions (based on research studies) was that the panelists tended to underestimate their uncertainty. However, in evaluating this assumption using various sensitivity analyses (Section 7.6), it became readily apparent that many of the panelists clearly did not underestimate their uncertainty. Additionally, the use of the study results, including the associated uncertainties, in any specific application is beyond the scope of the study. If the uncertainties are judged to be large in the context of any application, this simply reflects the uncertainties about loss-of-coolant accident (LOCA) frequencies as expressed by the panel responses. In applying the study results, it is up to the analyst to judge the extent to which the results are useful for their application. This point was clearly made in the last paragraphs of both the Executive Summary and the Conclusions (Section 9) of NUREG-1829.

Comment Number: NGC2
Comment: A distinction needs to be made between something being "state-of-the-art" or "the best that it can be", and "good enough for regulatory decision-making." The elicitation report does not address (directly or indirectly) the acceptability of using the results from the elicitation in the regulatory arena. Given its limitations, its use should be justified directly.

Response: As stated in the response to NGC1, the application of the study results, including the associated uncertainties, is beyond the scope of the study. If the uncertainties are judged to be large in the context of any application, this simply reflects the uncertainties about LOCA frequencies as expressed by the panel responses. In applying the study results, it is up to the analyst to judge the extent to which the results are useful for their application. This point was clearly made in the last paragraphs of both the Executive Summary and the Conclusions (Section 9) of NUREG-1829.

Comment Number: NGC3
Comment: Frankly, what we have in the elicitation is a series of informed, but at best, "best guesses" from knowledgeable experts in an area with essentially no experience data for LOCAs (other than SGTRs) and limited physical models. The elicitation report should justify why it is appropriate to manipulate these "best guesses" as if they were data drawn from sample spaces that exhibit a particularly fortunate statistical distribution (i.e., lognormal). Just as some people knowingly or unknowingly attribute more "truth" to a computer code result that displays a lot of decimal places, so some would assume that by statistically manipulating "best guess" answers, the inherent value, correctness, and fidelity of the guesses themselves increases.

Response: This comment implies that the estimated LOCA frequencies are merely the result of statistical manipulations of ”best guesses“ supplied by the panelists and bear no necessary relation to the true LOCA frequencies. To the contrary, the estimated LOCA frequencies are the result of an expert elicitation process, a well-established technique. The justification for choosing the expert elicitation process is clearly laid out in Section 1.5. It is used when neither operating experience nor physical models are sufficient to provide the technical basis necessary to answer the questions posed. This is clearly the case with evaluating generic LOCA frequency estimates.

The expert elicitation process has been used on a number of occasions to evaluate technical issues related to nuclear safety (Section 1.5). The process used in the study is an adaptation of the formal expert judgment processes used in NUREG-1150 to estimate core damage frequencies and in NUREG/CR-5411 to assess the performance of radioactive waste repositories (Section 3, p. 3-1) to site just a few examples. Because of the precedents created by these previous studies, no additional justification for using an expert elicitation process was deemed to be necessary and was not provided by NUREG-1829. It is also worth noting that the approach used has been generally endorsed by the Nuclear Regulatory Commission's (NRC's) Advisory Committee on Reactor Safeguards (Letter from M.V. Bonaca to N.J. Diaz dated April, 27, 2004, SECY-04-0037, Subject: "Issues Related to Proposed Rulemaking to Risk Inform Requirements Related to Large Break Loss-of-Coolant Accident (LOCA) Break Size and Plans for Rulemaking on LOCA with Coincident Loss-of-Offsite Power", Agencywide Document Access and Management System (ADAMS) Accession Number ML041200655).

Comment Number: NGC4
Comment: Fracture mechanics, and in particular probabilistic fracture mechanics, does not do a good job of predicting LOCA frequencies. The estimates by various codes and code users may vary by 4 to 6 orders of magnitude. There is no agreed upon physical model to capture the potential for future LOCAs.

Response: This comment reflects one of the principal reasons that expert elicitation was chosen to determine LOCA frequencies instead of relying solely on physical models. One of the principal codes used to evaluate intergranular stress corrosion cracking (IGSCC) issues in boiling water reactors (BWRs) is the PRAISE code which was developed in the early 1980's. Over the years, a number of enhancements have been made to PRAISE to extend its applicability and other codes have been developed to model failure frequencies for specific material degradation mechanisms. However, there is no single recognized code that models all significant LOCA precursor degradation mechanisms. Accordingly, it is to be expected that estimates produced by different codes and modelers may differ, sometimes dramatically. The lack of a single, generally-accepted physically-based code for predicting LOCA frequencies using only probabilistic fracture mechanics (PFM) is a reflection of the current scientific uncertainty about this subject. The NRC, Office of Nuclear Regulatory Research, has been sponsoring efforts to develop the next generation PFM codes to achieve this objective.

The limitations of existing PFM codes were well-known to the panelists during the elicitation process. One objective for supplementing the operating experience-based base case results with results obtained using two common PFM codes was to characterize systemic differences between the PFM codes for the same nominal input conditions. Because of these differences, almost all of the panel members used one of the two operating experience base cases for anchoring their elicitation responses. The PFM code base case results were used by some panelists to provide relative trends (not absolute values) for extrapolating current day LOCA estimates into the future and for extrapolating small break (SB) LOCA estimates, which are closely tied to operating experience (e.g., steam generator tube ruptures [SGTR] and through-wall cracking frequency), to larger-size LOCA estimates.

See the responses to NGC5, N3-1, N4-1, and N4-11 plus Comment 4-1 from Appendix M of the revised NUREG for additional information related to the use of PFM within this elicitation.

Comment Number: NGC5
Comment: For PFM codes, differences in assumptions and modeling between codes "can result in tremendous output variability in the pipe rupture probabilities." It is unclear which, if any, of the

assumptions are the appropriate ones to use. One single assumption difference between codes resulted "in up to a five order of magnitude difference in the conditional break probability predictions." The elicitation reports that "(w)hile the codes are complex, many simplifying assumptions are necessary and the model of actual plant conditions is, at best, approximate." It is not clear how closely the results of PFM reflect reality of the risk involved.

Response: Probabilistic fracture mechanics codes were used for two purposes during the elicitation: for providing base case estimates and for conducting sensitivity analyses to inform trend assessments. While the PFM and the operational experience predictions did not always readily agree with the base case analyses, this simply reflects the current uncertainty in the state of knowledge. The elicitation reported in NUREG-1829 did not attempt to resolve these discrepancies; rather, the elicitation's goal was to reflect the scientific uncertainty in the technical community.

This uncertainty was clearly evident to individual panel member and was used to inform the selection of their base case results for anchoring their elicitation responses and in providing uncertainty estimates. Generally, the panelists chose to anchor their responses to base case results based on operating experience instead of the PFM results. Therefore, the differences in absolute frequencies between the operating experience base cases and the PFM base cases are not very important in assessing variability among individual elicitation responses.

As noted in the response to NGC4, some panelists did use PFM for conducting trending analyses to provide a basis for extrapolating current day, small break (Category 1) LOCA frequencies to both future year (40 or 60 year) and larger-size LOCA frequency estimates. The PFM assumptions are less critical for assessing trends than in determining absolute frequencies. Hence, the commenter's concern that "…it is unclear which, if any, of the [modeling] assumptions are the appropriate ones to use." is mollified.

Key elements of this response were incorporated into Section 3.5 and 4.2 of the revised NUREG. Also, see the responses to NGC4, N3-1, N4-1, and N4-11 plus Comment 4-1 from Appendix M of the revised NUREG for additional information related to the use of PFM within this elicitation.

Comment Number: NGC6
Comment: It has been stated by RES that we do not know what new cracking phenomena will occur in the future, but we have noticed a trend that new phenomena tend to appear about every 7 years.

Response: A historical evaluation of operational experience indicates that precursors corresponding to generic nuclear reactor material degradation issues have occurred with an average periodicity of approximately seven years. Furthermore, operating experience indicates that both the NRC and the industry have aggressively responded to new degradation mechanisms when they arise. A good example is the nuclear community (NRC/industry) response to identifying and mitigating IGSCC in BWR plants approximately 25 years ago. The panelists were asked to assess the possibility that future degradation mechanisms will continue to appear and to assess the response and effectiveness of mitigation measures by both the NRC and the nuclear industry.

Comment Number: NGC7
Comment: Although the PFM models are attractive because they could be used to try to predict future piping degradation effects, the potential for new degradation mechanisms (e.g. something like PWSCC) was not evaluated and future degradation mechanisms are assumed to be handled swiftly and in time to prevent any increases in LOCA frequencies.

Response: As indicated to the responses to NGC4 and NGC5, PFM models were used by several panelists to predict relative failure trends related to specific degradation mechanisms for providing a basis for their elicitation responses. This is pointed out in the Executive Summary of NUREG-1829 where it states, "In this study, LOCA frequency estimates have been calculated using an expert elicitation process to consolidate operating experience and insights from PFM studies with knowledge of plant design, operation, and material performance."

Current operating experience contains information related to response rate and the effectiveness of mitigation measures in addressing several degradation mechanisms (e.g., IGSCC or flow accelerated corrosion [FAC]). This history provided a basis for assessing the impact of future degradation mechanisms which may not be apparent in the current operational-experience database. No panelist assumed that future degradation mechanisms will "…be handled swiftly and in time to prevent any increases in LOCA frequencies" as stated in the comment.

Rather, a number of the panelists indicated that if a new degradation mechanism arises in the future, the LOCA frequency may temporarily increase until the industry has developed an effective mitigation strategy. Only after this mitigation strategy has been fully implemented will the LOCA frequencies decrease to values commensurate with those prior to the manifestation of the new degradation mechanism. Historical operating experience and PFM insights provided a basis for assessing what the increase may be, and how long it may take to mitigate the risk.

Sections 6.3.2 and 6.3.4 of NUREG-1829 has been modified to incorporate principal elements of this response as well as additional information on the emergence of future degradation mechanisms and the role of mitigation strategies.

Comment Number: NGC8
Comment: Specific plants have applied single or multiple mitigation strategies to help prevent pipe cracking and LOCAs. This variation in mitigation is not well captured in the evaluation.

Response: It is true that plants have used different mitigation strategies to decrease the LOCA risk associated with specific degradation mechanisms. The quantitative effectiveness of these strategies will be a function of the effectiveness of the baseline strategy as well as plant-specific variables that modify the effectiveness (i.e., environment, residual stress, operating stress, plant configuration, etc.). As such, mitigation effectiveness is expected to be a plant-specific variable. Operational experience contains information on the general effectiveness of industry mitigation measures and also contains information related to particularly ineffective strategies employed at specific plants. More subtle differences are not apparent in operational experience. However, this situation is no different than the effect of other plant-specific variables, i.e., those related to materials, geometries, operating environments, etc. Their effects are not manifested in the operational-experience database unless they are significant enough to result in either a failure or a precursor event (e.g., cracking).

Furthermore, as stated in Section 2 of NUREG-1829, "… the elicitation focused on developing generic, or average, estimates for the commercial fleet and uncertainty bounds on these generic estimates. The BWR and pressurized water reactor (PWR) LOCA frequency estimates were not partitioned further to describe differences related to design class, vendor, or specific plant operating characteristics. This approach is consistent with prior LOCA frequency studies that did not consider plant-specific differences in estimating LOCA frequencies for use in probabilistic risk assessment (PRA) modeling. However, the elicitation panelists were instructed to consider broad differences among plants related to important variables (i.e., plant system, material, geometry, degradation mechanism, loading, mitigation/maintenance) in determining both the generic LOCA frequency estimates and especially the uncertainty bounds. It is the broad differences in these important variables that are most important to passive system failure and there is generally sufficient commonality among plants to make such a generic assessment meaningful." Therefore, the elicitation estimates are expected to be representative of most plants.

However, it is also understood that unique plant features and safety culture may also influence LOCA frequencies and plant-specific estimates may fall outside the generic fleet estimates. However, accident frequency increases stemming from deficient safety practices are expected to be identified and mitigated through regulatory oversight policies and procedures.

Section 2 of NUREG-1829 has been clarified based on this comment

Comment Number: NGC9
Comment: The report has an inadequate discussion of non-piping system failure modes. It is impossible to judge the adequacy of the work performed.

Response: In general, the non-piping elicitation structure, available background information and base case development was completely analogous to the corresponding evaluation of piping system contributions to the LOCA frequencies. When the elicitation started, less operating experience data had been collected specifically for non-piping system failures. Therefore, a non-piping failure precursor database was developed from licensee event reports (LERs) and provided to the panelists to correct this initial deficiency. This database is analogous to the piping precursor failure database that was provided, although it is not as comprehensive. Details of the non-piping precursor database are provided in Appendix H of NUREG-1829.

Additionally, several panelists had experience evaluating specific non-piping degradation and failure mechanisms that have occurred in service. These evaluations were used to develop specific non-piping base cases that were available for anchoring subsequent elicitation responses in a completely analogous manner to the piping base cases. These non-piping base case evaluations included: SGTR frequency determination, current pressurized thermal shock (PTS) analysis, BWR vessel rupture due to beltline cracking, BWR feedwater nozzle thermal fatigue cracking, and PWR control rod drive mechanism (CRDM) failure due to pressurized water stress corrosion cracking (PWSCC). Information on the analysis details and results for each of these non-piping base cases is provided in Section 4.4 of NUREG-1829. Appendix I provides additional details and results for BWR vessel rupture, feedwater cracking, and PWR CRDM failure. It should also be noted that the piping base case evaluations were also available for anchoring non-piping elicitation responses. Some panelists choose these for anchoring in appropriate situations.

The elicitation structure for evaluation non-piping LOCA frequency contributions was also analogous to the structure chosen to estimate piping LOCA frequency contributions. Discussion of non-piping LOCA frequency contributions can be found in the minutes to the Elicitation Kick-Off meeting in Appendix B, pages B-32 through B-37. The specific failure modes (or scenarios) for the five non-piping components considered to contribute to the LOCA frequencies, i.e., reactor pressure vessel (RPV), pumps, valves (BWRs and PWRs) and pressurizer and steam generators (PWRs), can be found in Tables B.1.13 through B.1.17 of Appendix B.

There were only two notable differences between the piping and non-piping evaluations. The first difference is that several panelists performed separate base case evaluations for the defined piping scenarios while a single panelist performed each non-piping evaluation. As mentioned, the multiple piping evaluations were used to identify differences in the estimated LOCA frequencies associated with the base cases, based on the analysis approach and assumptions. It was not necessary to repeat this study for similar non-piping scenarios. The second difference is that some of the non-piping failure modes considered were distinct from important piping failure modes and did not lend themselves to classical modeling approaches. Some examples include common cause bolting failures resulting from maintenance that could lead to vessel head, pump bonnet, valve bonnet, or steam generator manway failures. It is for these types of failure modes that elicitation is most valuable.

It is also instructive to review the results to determine the importance that the elicitation panelists placed on non-piping contributions to the LOCA frequencies. The piping contributions to the BWR LOCA frequencies were more significant than the non-piping contributions for LOCA Categories 1–5 (Figure 7.6). The largest Category 6 LOCAs were governed by BWR vessel rupture, since no BWR piping system is large enough to sustain a Category 6 LOCA. The PWR LOCA frequencies (Figure 7.7) are dominated by the SGTR contribution for Category 1, while CRDM failures are the main contributor to Category 2 LOCAs. For PWR LOCA Categories 3 through 6, the piping and non-piping contributions are nearly the same. More importantly, the significant non-piping failure modes identified by the elicitation panelists (Table 6.1) were analogous to common piping failure modes and usually were directly related to a specific non-piping or piping base case scenario.

Key elements of this response have been incorporated into Section 3.5.2 of NUREG-1829.

Comment Number: NGC10
Comment: There are many LOCA frequency estimates provided in the report that are so low as to be unbelievable. Such low numbers only exhibit our lack of understanding of the limits of the models and the physical phenomena involved. When one uses frequencies with return periods millions or billions of years longer than those of massive meteor strikes on earth or disastrous volcanic eruptions that covered thousands of square miles with lava, it is clear that our limitations of knowledge drive the number so low. No one should believe frequencies orders of magnitude longer than the existence of the universe.

Response: This comment supposes that an analysis of an event which leads to a frequency which is so low that the occurrence of the event is incredible and concludes that the analysis is therefore unbelievable. However, the accuracy of an analysis that results in a very low frequency depends on the validity of the assumptions and modeling approach rather than the value of the frequency per se. An analogous situation is the following. Suppose someone purchases one ticket in each of three successive months, each having one chance in a million of winning. The probability of winning all three times is $(1E\text{-}6)^3 = 1E\text{-}18$. This is truly an incredible event, but the calculation of its probability is correct nonetheless. However, lower frequency estimates are typically more sensitive to the accuracy and variability of critical underlying modeling and assumptions because changes are more likely to have a proportionally bigger impact. Therefore, the only direct conclusion that can be drawn from an extremely low LOCA frequency estimate is that the occurrence of such an event is incredible based on the assumptions, input distributions, and model used for the prediction.

There are several extremely low frequencies (1E-10 per calendar year to 1E-20 per calendar year) provided in NUREG-1829. These low frequencies are most often associated with various PFM analyses. Several base case frequencies were calculated to be less than 1E-10 per calendar year as shown in Table 4.1 and Figures 4.2 and 4.3. The sensitivity analyses for the effect of hydro test loading (Figure 4.14; frequency < 1E-16 per calendar year), the effect of a safe shutdown earthquake (SSE) load (Figure 4.15; frequency < 1E-16 per calendar year), and the effect of the material's fracture toughness (Figure 4.16; frequency < 1E-13 per calendar year) also exhibit similarly low values. There are also several extremely low frequencies reported in Appendix F related to the piping base case results and additional sensitivity analyses.

As discussed, the validity of these extremely low estimates is primarily a function of the validity of the assumptions and the modeling approach. For instance, one of the PFM evaluations for PWR base case 1 (hot leg cracking due to PWSCC) only considered failures due to fatigue from preexisting flaws while the second considered both fatigue and PWSCC and allowed flaws to initiate during operation. There is additional discussion in Section 4.2 of NUREG-1829 for the differences among the base case results, and of causes for the extremely low frequencies calculated in certain evaluations. Section 4.3 contains similar discussion for the results of the base case sensitivity studies. The PFM assumptions, modeling inputs and limitations are clearly detailed in Section 3.5 and Appendices F and G. The assumptions, results, and implications from the various PFM analyses were also articulated to the panelists in several of the meetings, so that their elicitation responses were appropriately informed. Panelists generally did not consider these extremely low estimates (< 1E-10 per calendar year) to be representative of important LOCA frequency contributions for a specific system, except for possibly the very largest LOCA Categories (i.e., 5 and 6).

Key elements of this response were incorporated into Section 4.2 of the revised NUREG.

Comment Number: NGC11
Comment: Much of what is in the report is difficult to understand, even for someone with advanced degrees in mathematics and engineering.

Response: The subject matter of the report is complex. However, it is important to include the details of the methodology so that experts in the field can adequately judge the merits of the approach and the validity of the conclusions. Making the report more "reader friendly" by eliminating some of these very details may make it more readable but would diminish the level of documentation and value of the report. In writing the report, the authors tried to present the results so that a technically oriented non-expert could appreciate and understand the conclusions while including enough details so that an expert could judge the validity of the methodology.

Comments Related to Executive Summary of NUREG-1829

Comment Number: NES1
Comment: The expert elicitation process and results described in the draft NUREG were developed to support the selection of a new transition break size (TBS) LOCA in the proposed 10CFR50.46a rule. The draft NUREG aggregates the panelists' estimates of break frequencies using several different methods and each aggregation method yields a different break size versus cumulative frequency (break-frequency) curve. The draft NUREG elevates one particular aggregation methodology and its resulting break-frequency curve to a baseline estimate and thereby assigns a special status to this curve. Any baseline break-frequency curve will become the preferred curve that will be used whenever a LOCA frequency is needed in future calculations, including those used to support licensing action requests. Furthermore, subsequent deliberations on selecting a TBS for the proposed rule may rely more heavily on a single break-frequency curve than the current deliberations have. Selection of the single break-frequency curve would normally default to the "baseline" curve. The aggregation method used to develop the proposed baseline break-frequency curve in the draft NUREG yields substantially different result than aggregation methods selected by the staff to represent the baseline results of previous expert elicitations. NUREG-1150 (Severe Accident Risks) and NUREG-1488 (Seismic Hazard Curves) develop the baseline results by aggregating the panelists' estimates using the arithmetic mean. The draft NUREG develops its baseline results by aggregating the panelists' estimates using the geometric mean. The geometric mean always yields lower aggregate mean estimates and always results in tighter uncertainty bounds compared to the arithmetic mean. For example, selecting the break size that corresponds to 1E-5/year frequency yields 6 inches for the geometric mean versus 12 inches for the arithmetic mean for boiling-water reactors (BWRs). For pressurized-water reactors (PWRs) the geometric mean yields 3 inches while the arithmetic mean yields 6 inches. Elevation of a single break-frequency curve to the baseline curve as proposed by the draft NUREG is not consistent with the way the curves were used to support the development of the

TBS for the rulemaking. Instead of using one curve, a frequency was selected and the range of break sizes at that frequency but from different aggregation methods was identified. Qualitative considerations including, in part, actual piping system design and operating experience were used to characterize the break spectrum for PWRs and BWRs. The characteristics of this break spectrum were compared to the range of break sizes for the selected frequency from the alternative aggregation techniques. A TBS was chosen that depends on the characteristics of the piping systems at each plant and is consistent with the range of break sizes from the break-frequency curves and the uncertainty in the expert elicitation process. There is a consensus within the NRR and RES staff that the development of the break-frequency curves as described in the draft NUREG is sufficient to support the proposed rulemaking as described in SECY-05-0003. There is, however, no consensus within the staff on which aggregation technique should be used to support a baseline LOCA 2 break-frequency curve, i.e., the curve that would represent the best estimate break frequency curve when a single curve must be selected. Derivation of the correct aggregation technique is not possible. The geometric mean supports the goal of developing a consensus group mean value by reducing the impact of wide differences in the individual estimates. The arithmetic mean supports the goal of preserving the full range of uncertainty in the mean value by assuming that all the panelists' estimates are based on equally credible models. The NUREG should make it clear that the break-frequency curve presented as the baseline estimate has not been endorsed as the best estimate LOCA frequency curve by the staff.

Response: This comment implies that, by not following the precedent of using the arithmetic mean for aggregating the panelists' estimates as was done in the cited studies, the draft NUREG uses an inappropriate aggregation scheme. However, as noted in the response to Comment N5-3 and in Section 5.6.4.1 of the NUREG, the geometric mean is an appropriate aggregation measure for calculating consensus-type group estimates, especially when, as in this elicitation, the estimated frequencies differ by several orders of magnitude. Furthermore, as discussed in the response to Comment N5-4 and Section 5.6.4.1, there is a precedent, based on NRC-sponsored work, for using the median (well-approximated by the geometric mean for the NUREG results) to aggregate expert judgments.

The commenter is correct that arithmetic-mean aggregation will always result in higher estimates. Therefore, sensitivity analyses were conducted to evaluate the effect of using mixture-distribution and arithmetic-mean aggregation to calculate group estimates. In this study, these alternative aggregation methods can lead to significantly higher mean and 95th percentile estimates than those obtained using geometric-mean aggregation. The biggest differences occurred when the estimates from one or two panelists were significantly higher than the remainder of the group. Because of these discrepancies, NUREG-1829 does not recommend any particular LOCA frequency estimates for all risk-informed applications.

The last paragraphs of the Executive Summary and the Section 9 (Conclusions) have been modified as follows to address this commenter's concern:

> Because the alternative aggregation methods can lead to significantly different results, a particular set of LOCA frequency estimates is not generically recommended for all risk-informed applications. The purposes and context of the application must be considered when determining the appropriateness of any set of elicitation results. In particular, during the selection of the BWR and PWR transition break sizes for the proposed 10CFR50.46a rule making, the NRC staff considered the totality of the results from the sensitivity studies rather than only the summary estimates from this study The NRC anticipates that a similar approach will be used in selecting appropriate replacement frequencies for NUREG/CR-5750 estimates and for other applications where frequencies for break sizes other than those in NUREG/CR-5750 are required. While the lack of clear application guidance places an additional burden on the users of the study results,

those users are in the best position to judge which study results are most appropriate to consider for their particular applications.

Additional information on the differences between the geometric mean aggregated results and other alternative schemes is contained in Section 7.6.4.

Comment Number: NES2
Comment: The break frequency estimates are based on degradation related failure frequencies. However, the report acknowledges that there are other contributors to the overall failure frequencies. In particular, seismic failure frequencies may have a dominant effect in the larger size range when considered with various levels of degradation. A description of the recent study sponsored by the Office of Nuclear Regulatory Research regarding seismic failure frequencies should be provided and what is involved in making such estimates. Similarly, in the smaller size range, the effect of active system LOCAs may be significant. It should be emphasized that the user of the NUREG should include all other significant contributors to failure frequencies.

Response: The initial application of NUREG-1829 was to select the TBS for use in the risk-informed revision of 10 CFR 50.46. Considerations other than just the NUREG-1829 results were used in this selection. As stated on Page 67603 of the Federal Register Notice (Risk-Informed Changes to Loss-of-Coolant Accident Technical Requirements; Proposed Rule),

> *A baseline TBS (transition break size) was established using these break frequencies (those from the expert elicitation) as a starting point. This baseline TBS was then adjusted to account for other significant contributing factors that were not explicitly addressed in the expert elicitation process. The following three-step process was used by the NRC in establishing the TBS.*
> *(1) Break sizes for each reactor type (i.e., PWR and BWR) were selected that corresponded to a break frequency of once per 100,000 reactor years (i.e., 1.0E-5 per reactor year) from the expert elicitation process.*
> *(2) The NRC then considered uncertainty in the elicitation process, other potential mechanisms that could cause pipe failure that were not explicitly considered in the expert elicitation process, and the high susceptibility to rupture/failure of specific piping in the RCS.*
> *(3) The NRC adjusted the TBS upwards to account for these factors.*

As part of the study to examine the effect of seismic loading on the selection of the TBS (Reference: Chokshi, N.C., Shaukat, S.K., Hiser A.L., DeGrassi, G., Wilkowski, G., Olson, R., and Johnson, J.J., "Seismic Considerations For the Transition Break Size, " NUREG-1903, U.S. Nuclear Regulatory Commission, February 2008), both unflawed, flawed, and indirect seismic piping failure analyses were conducted in order to ascertain the magnitude of any potential adjustments to the baseline TBS established as part of the expert elicitation process. The intent of this study was not to perform bounding analyses that will encompass all potential variations, including site-to-site variability in the seismic hazard curve. Rather, the purpose was to determine the relative importance of seismic effects on the proposed TBS and to provide information on key considerations in order to elicit comments and information germane to this issue during the public review and comment period.

The principal findings from this study are that the critical flaws associated with the stresses induced by seismic events of 10^{-5} and 10^{-6} annual probability of exceedance are large, and coupled with other mitigating factors, the probabilities of pipe breaks larger than the TBS are likely to be less than 10^{-5} per year. Similarly, for the cases studied, the probabilities of indirect failures of large RCS piping systems are less than 10^{-5} per year.

As a result of this comment, the NRC report on *Seismic Considerations for the Transition Break Size (NUREG-1903)* is now referenced in Section 2 of NUREG-1829. In addition, a summary of the seismic LOCA analysis and results is provided in the Executive Summary and in Section 7.2 of the report. Additionally, Section 2 has a clear discussion of the scope and assumptions underlying the elicitation LOCA frequency estimates. In particular, the NUREG states that "…rare event loading from seismic, severe water hammer, and other sources was also not considered in this generic evaluation because of their strong dependency on plant specific factors."

The NUREG, in Section 2, also emphasizes that users should consider all other significant contributors to LOCA frequencies, and not just the contributions from the passive degradation mechanisms addressed in expert elicitation, if comprehensive LOCA frequency estimates, spanning all break sizes are desired. It has also been emphasized in Section 1.4 that contributions from active mechanisms to the overall failure frequencies should also be considered for developing total LOCA frequency estimates. As indicated in Section 1.4, the expectation is that active system failures can be estimated directly from operating experience as is customary.

Comments Related to Section 1 of NUREG-1829

Comment Number: N1-1
Comment: The discussion of seismic-induced LOCAs is misleading in the Background discussion of the elicitation report. The numbers presented appear to represent the conditional probabilities of large LOCAs given SSE loads. The frequency of earthquakes greater than an SSE could be in the 10^{-4} to 10^{-3} per year range for some plants and in the 10^{-5} per year for many plants. Since seismic hazard curves have such steep slopes, a small change in return frequency can result in a very large change in acceleration levels and therefore seismic loads.

Response: The numbers presented are drawn from the Lawrence Livermore National Laboratory (LLNL) study conducted in the 1980s which estimated the frequency of a reactor coolant system (RCS) double-ended guillotine break (DEGB) for each of the four U.S. reactor vendor designs [i.e., Westinghouse, Combustion Engineering (CE), General Electric (GE), and Babcock & Wilcox (B&W)]. The objective of this program was to estimate the frequency of a DEGB attributable to (1) pipe fracture directly caused by the growth of cracks at welded joints, termed "direct DEGB," and (2) pipe rupture indirectly caused by failure of major components or component supports as a result of an earthquake, termed "indirect DEGB." These studies are documented in a variety of NUREG-series reports as indicated in NUREG-1829.

Overall, the LLNL study concluded that the frequency of a direct DEGB in RCS piping is very low for both PWR and BWR plants. LLNL further concluded that the probability of a leak or direct DEGB in the RCS piping is negligibly affected by earthquakes, to the extent that direct DEGB and earthquake can be considered independent random events. The LLNL study also concluded that indirect causes are clearly the more dominant mechanism leading to a DEGB in RCS piping. The probability of an indirect DEGB is a strong function of seismic hazard, but is nonetheless low even when earthquakes significantly greater than the SSE are considered. Table 1.2 in NUREG-1829 also summarizes ranges of results for both direct and indirect DEGB failure for the sample PWR plants of each type that were evaluated.

Comment Number: N1-2
Comment: The fact that elicitations have been used in the past to try and bring insight into areas with little data does not mean that the elicitations are particularly accurate, complete, or necessarily appropriate to use in the regulatory decision-making arena.

Response: As stated in the NUREG, there are many precedents for using elicitation to support regulatory decision-making, including studies to evaluate reactor risk (NUREG-1150), to develop seismic hazard curves (NUREG/CR-6372), and to assess the performance of radioactive waste repositories (NUREG/CR-5411). The purposes and context of the regulatory application must be carefully considered when determining the appropriateness of any set of elicitation results and the effect of uncertainties in the results should be also be evaluated. In particular, the BWR and PWR TBSs for the proposed 10CFR50.46a rule making, were selected by NRC staff after considering the effects of using different analysis methods (i.e., the sensitivity studies on the TBS). Additional information justifying the use of expert elicitation for this study and describing prior regulatory uses of elicitation are provided in Sections 1 and 5 of the revised NUREG. Also, see the response to Comments NGC1 and NGC2 for similar discussion.

Comment Number: N1-3
Comment: "The elicitation focused on developing generic or average values." So it is not clear to what extent its conclusions are applicable to outlier plants, older plants, plants with safety culture problems, plants that had poor QA/QC, or in general any plant that strays from the norm. Panelists were instructed "not to consider differences that only exist at a few plants." The results therefore may neither describe the true expected LOCA frequencies for some or even many plants, and the upper bound estimates reported may not truly bound the real expected frequencies at outlier plants, since they were not even considered in the evaluation. Only the most common piping materials and welds were considered.

Response: It is correct that the NUREG focused on generic or average values (see Section 2). However, the elicitation panelists were instructed to consider broad differences among plants related to important variables (i.e., plant system, material, geometry, degradation mechanism, loading, mitigation/maintenance) in determining both the generic LOCA frequency estimates and especially the uncertainty bounds. It is the broad differences in these important variables that are most important to passive system failure and there is generally sufficient commonality among plants to make such a generic assessment meaningful. However, while broad plant variability was considered, the study results do not identify its specific effects.

Additionally, panelists were specifically instructed not to consider differences that only exist at a few plants. For instance, if a particular plant vendor design was judged by a panelist to be more LOCA-sensitive and encompasses, say, 20 PWRs, then the panelist was expected to consider this in his estimates as well as in the uncertainty bounds.. However, if the same plant design only applies to one or two plants, then this should not have been considered in the estimates.

It is understood that unique plant features and safety culture may also influence LOCA frequencies and plant-specific estimates may fall outside the generic fleet estimates. However, accident frequency increases stemming from deficient safety practices are expected to be identified and mitigated through regulatory oversight policies and procedures. Also, the staff was directed to provide realistically conservative LOCA frequency estimations (staff requirements memorandum (SRM) to SECY-02—57, ML030910476), and not bounding values associated with one or two plants. This approach is also consistent with prior LOCA frequency studies that did not consider plant-specific differences in estimating LOCA frequencies for use in PRA modeling. The objective of the NUREG-1829 study was to develop results in a similar context. Therefore, the results satisfy both the Commission direction and the historical development and use of these LOCA frequency estimates. Section 2 has been modified to address this comment.

With regard to the comment that “Only the most common piping materials and welds were considered”, the panelists focused their attention on those materials and welds that, based on operating experience, are most susceptible to potential degradation mechanisms, e.g., stainless steel welds subjected to IGSCC and dissimilar metal welds subjected to PWSCC. Specifically, each panelist was asked to identify specific combinations of the five variable classes (material, geometry, degradation mechanism, loading, mitigation/maintenance) whose total contribution to LOCA frequency for a given system is at least 80% of the contributions of all possible variable combinations. Additionally, the contributions of variable combinations not specifically identified were not ignored. Their contributions were accounted for by the LOCA frequencies as adjusted for the total contribution factor. For example, if a panelist said that the identified variable combinations accounted for 80% of the total contribution to LOCA frequency, then that panelist's LOCA frequency was multiplied by 1.25 = 1/0.80. (See Section 5.2 and Appendix J, Section J.2.3.)

Comment Number: N1-4
Comment: The elicitation assumes that "once degradation mechanisms are discovered [in the future] and assessed, it is expected that mitigation measures will be developed and applied in order to decrease the passive system degradation rate to historical levels, or lower." There is no basis for this optimism that **any and every** future mechanism will be discovered in time and mitigated in such a manner so that it will not affect the assumed LOCA frequency distributions of today. It is worth noting that utilities today are struggling with the loss of expertise to read and diagnose pipe defects that can be revealed by various detection techniques.

Response: This is not an assumption of the elicitation at all. The effect of future mechanisms on the LOCA frequency estimates was something that was required for each panelist to provide in their elicitation responses. Some panelists did agree with the stipulation that mitigation measures will be developed and applied to counter near-term frequency increases due to age-related degradation. These panelists typically cited past experience as the underlying basis for this contention. Historically, these panelists argue, each time a new degradation mechanism has arisen, the NRC and industry have developed mitigation strategies to address that mechanism. However, that does not mean that LOCA frequencies do not increase at all according to these panelists. These panelists typically predicted a short term increase in frequency estimates until the mitigation strategies were universally adopted. At this point, frequencies would decrease back to historical levels. Not every panelist adopted this view and subsequently provided estimates which increased with continued operation. However, this view was shared by a majority of the panelists which explains why the quantitative LOCA frequency estimates are only moderately higher in the upcoming 15 and 35 years compared with the current-day estimates (Section 7.4).

The effect of lost expertise for diagnosing piping defects was considered in the elicitation along with a related concern that as plants approach the end of their current licensing periods, the asset management philosophy adopted by individual utilities may not lend itself to continuing maintenance practices necessary to mitigate degradation.. These types of issues were addressed within the context of the effects of safety culture on future LOCA frequencies. Some panelists did consider that these, and other, issues could increase the LOCA frequencies with time. However, compensating factors (e.g., additional industry experience, deregulation, improved technologies, etc.) act to offset LOCA frequency increases. On balance, the elicitation panel responses indicate that future safety culture changes are not expected to significantly alter the LOCA frequency estimates. See Sections 6.2 and 7.1 for more additional discussion on the effects of safety culture.

Key elements of this response have been incorporated into Executive Summary and Sections 1.3, 6.3.2 and 7.4 of the revised NUREG.

Comments Related to Section 2 of NUREG-1829

Comment Number: N2-1
Comment: The NUREG makes it clear that the LOCA frequencies set forth are intended to be averages across the industry. The TBS in the currently proposed 50.46 rule is developed based on these curves and is used for all BWRs and all PWRs. Normally; bounding values are used when a parameter is to be applied across the industry and not the average value. Alternatively, the characteristics associated with the average value are provided and the individual plants must ensure they lie within the envelope. The difficulty associated with applying average values to the entire industry is recognized in the NUREG. Page 2-1, of the NUREG states:

> The elicitation focused on developing generic, or average, values for the commercial fleet and uncertainty bounds of this generic average. The BWR and PWR LOCA frequencies have not been partitioned further to describe differences related to design class, vendor, or specific plant operating characteristics. These features can influence LOCA frequencies and it is expected that frequency distributions for specific plants differ from these generic distributions.

One example of the use of an average parameter causing an uncertain impact on the proposed application of the results to the industry is the conversion of the flow rates in gallons per minute to break size. On page 2-1, the NUREG states: "Simple correlations were developed to relate the rupture size to the expected flow rate required for the emergency core cooling system make-up system." The "simple correlations" are provided for BWR (both steam and liquid) and PWR (liquid only). Later, in Section 3.7, it further says: "Application of the elicitation LOCA frequency results may require plant specific calculations to evaluate break flow rate history and required mitigation response for assumed break locations and sizes."

The generic LOCA frequencies derived in the NUREG may not be representative of a given plant, are not bounding estimates, and there does not appear to be an attempt to define the applicable envelope.

Response: See the response to Comment N1-3 which is partly reconstructed here. It is correct that the NUREG focused on generic or average values (see Section 2). Elicitation panelists were instructed to consider broad differences among plants related to important variables (i.e., plant system, material, geometry, degradation mechanism, loading, mitigation/maintenance) in determining both the generic LOCA frequency estimates and especially the uncertainty bounds. It is the broad differences in these important variables that are most important to passive system failure and there is generally sufficient commonality among plants to make such a generic assessment meaningful. However, while broad plant variability was considered, the study results do not identify its specific effects.

Additionally, panelists were specifically instructed not to consider differences that only exist at a few plants. For instance, if a particular plant vendor design was judged by a panelist to be more LOCA-sensitive than other designs and encompasses, say, 20 PWRs, then the panelist was expected to consider this in his estimates as well as in the uncertainty bounds.. However, if the same plant design only applies to one or two plants, then this should not have been considered in the estimates.

It is understood that unique plant features and safety culture may also influence LOCA frequencies and plant-specific estimates may fall outside the generic fleet estimates. However, accident frequency increases stemming from deficient safety practices are expected to be identified and mitigated through regulatory oversight policies and procedures. Also, the staff was directed to provide realistically conservative LOCA frequency estimations (SRM to SECY-02—57, ML030910476), and not bounding values associated with one or two plants. This approach is also consistent with prior LOCA frequency studies that did not consider plant-specific differences in estimating LOCA frequencies for use in PRA modeling. The objective of the NUREG-1829 study was to develop results in a similar context. Therefore, the results satisfy both the Commission direction and the historical development and use of these LOCA frequency estimates. Section 2 of the NUREG has been modified to address this comment.

Comments Related to Section 3 of NUREG-1829

Comment Number: N3-1
Comment: It is not apparent from the results of this expert elicitation that PFM is a sufficiently robust evaluation tool to be used for making regulatory decisions of the magnitude and scope of modifications to 10CFR50.46. The elicitation was performed at Commission behest expressly to try and bring PFM results and insights closer to those of service history-bases analyses. It appears the elicitation has failed to do that.

Response: The elicitation was designed to allow the panelists to use their best judgment in making their LOCA frequency estimates, after being made aware of various possible approaches, uncertainties associated with each approach, and differences in results among competing approaches. This was the intent of the base case analyses that were performed and they, in the authors' opinion, achieved that objective. The staff direction from the Commission was to use expert elicitation to converge results based solely on operating experience and PFM. That was the focus of the elicitation. All panel members were knowledgeable of both the operating experience and the PFM of not just the base cases, but also differences stemming from their use to analyze similar problems in the past. The extent to which operating experience and PFM insights was used by each panelist is reflected in their approach to responding to the elicitation questions. The goal in developing the elicitation framework was to accurately reflect the panelists' opinions without introducing a preferred approach. See the responses to NGC4, NGC5, N4-1, and N4-11 plus Comment 4-1 from Appendix M for additional information related to the use of PFM within this elicitation.

Comment Number: N3-2
Comment: The elicitation was performed for plants subject to only nominal conditions. No unusual transients or accidents were assumed to occur. Plant operation was assumed to be the same as it is today. That is, there were no assumed power upgrades, there were no changes in water chemistry practices, there were no unusual transient assumed to occur to challenge the plant.

Response: The effect of transients was considered in the elicitation although these were limited to transients expected over the operating life of a plant (i.e., frequency greater than 1/60 per year). This assumption was necessary so that the past operating experience was applicable and could be used as a basis for elicitation responses. Significant changes in operating performance may invalidate the operating experience which is why this restriction was necessary. This assumption is clearly stated in the Executive Summary and scope of the report (Section 2).

Additionally, the effects of changes in the plant operational environment (i.e., loss-of-knowledge, deregulation, decommissioning, etc) were discussed with the panel members during their individual elicitations as part of their evaluation of the effects of safety culture on LOCA frequencies. As a result of these discussions it was concluded that safety culture effects on future generic LOCA frequencies are expected to be minimal. It is understood that safety culture may also influence LOCA frequencies and plant-specific estimates may fall outside the generic fleet estimates. However, the elicitation panelists generally expect that accident frequency increases stemming from deficient safety practices will continue to be identified and mitigated through regulatory oversight policies and procedures. See Section 7.1 of NUREG-1829 for additional discussion on safety culture effects.

Additionally, a separate study (Reference: Chokshi, N.C., Shaukat, S.K., Hiser A.L., DeGrassi, G., Wilkowski, G., Olson, R., and Johnson, J.J., "Seismic Considerations For the Transition Break Size, " NUREG-1903, U.S. Nuclear Regulatory Commission, February 2008) was conducted to examine the impact of unusual seismic events on the TBS. These transients typically generate the most additional failure risk. The key findings from this study for unflawed piping are that seismic-induced failure probabilities of unflawed piping are significantly lower than the frequency of 10^{-5} per year used as a basis to establish the TBS. For flawed piping, this study found that the critical flaws associated with the stresses induced by seismic events of 10^{-5} and 10^{-6} annual probability of exceedance are large, and coupled with other mitigation factors, the probabilities of pipe breaks larger than the TBS are likely to be less than 10^{-5} per year."

Key elements of this response have been incorporated into the Executive Summary and Sections 2, 7.1 and 7.2 of the revised NUREG.

Comment Number: N3-3
Comment: The panel participants (materials experts) did not directly provide LOCA frequencies to the facilitation team, but instead provided ratios that reflected how they believed pipe break flow rates would change compared to base cases. It is unclear to what extent the panelists understood the implications of their "best guesses" as to how the ratios would change.

Response: The basic premise behind the elicitation philosophy (see Section 3.8.1 of NUREG-1829) was that "an assessment of the relative likelihoods for comparable LOCA frequency contributions is a more natural reflection of knowledge and experience than is an assessment of the absolute frequencies for each contributor." As stated in Section 3.8.1 of NUREG-1829;

> *A relative assessment allows a direct comparison of the effects on LOCA frequencies resulting from specific combinations of materials, loading characteristics, aging mechanisms, and mitigation procedures for all the LOCA-sensitive components. Assessments of these interactions and comparisons of these relationships best match the panel expertise. An assessment of absolute LOCA frequencies would not only require consideration of these relationships, but would also require estimation of the frequencies associated with each postulated set of conditions. However, all of the conditions rarely occur, and there is little supporting data for the members to assess the absolute frequencies. Consequently, assessing relative frequencies should lead to more accurate results.*

As part of the kick-off meeting for this elicitation study the panel members were trained in this approach and saw first hand the advantages of making relative estimates (i.e., ratios) instead of absolute estimates by comparing their responses to almanac type questions (for which known answers were available) as described in Section 3.3.2. The results of this training exercise demonstrated that the panel provided more accurate responses using relative ratios rather than absolute numbers. This finding is consistent with the assumption behind the basic structure of this elicitation, namely, that relative values are easier to assess than absolute values.

While this framework was used for the elicitation, the elicitation panelists were also provided with their analyzed responses so that they could understand the implications of their individual elicitation responses and how they contributed to their estimated LOCA frequencies. The panelists then had the opportunity to modify their responses if the they believed that the frequencies did not adequately reflect their elicited opinions. However, none of the panelists chose to modify their responses. See Section 3 of NUREG-1829 for additional information.

Comment Number: N3-4
Comment: Panel members believe that the effects of safety culture can significantly affect LOCA frequencies at a specific plant. This was not factored into the elicitation.

Response: As discussed in Section 6.2 of NUREG-1829, many panelists expressed the opinion that safety culture is a highly plant-specific factor. Most panelists believe that the industry is generally acting in a consistent, safety conscious manner. However, individual plants can deviate markedly from general industry philosophies and practices. The vessel head degradation at Davis-Besse was a commonly cited example. Panelists argued that the LOCA frequencies at these less safety conscious plants could be greatly elevated compared with the remaining population. However, this trend would not affect either the median LOCA frequency or safety culture estimates, but would more likely affect the upper bounds of both distributions. Therefore, the upper bound estimates do incorporate (as indicated in the responses to Comments N1-3 and N2-1) broad plant-specific variability.

The upper bounds do not, however, account for egregious safety culture deficiencies that may exist in any single plant. It is understood that unique plant features and safety culture may also influence LOCA frequencies and plant-specific estimates may fall outside the generic fleet estimates. However, the elicitation panelists generally expect that accident frequency increases stemming from deficient safety practices will continue to be identified and mitigated through regulatory oversight policies and procedures. See Section 7.1 of NUREG-1829 for additional discussion on safety culture effects.

Key elements of this response have been incorporated into the Executive Summary and Sections 2 and 7.1 of the revised NUREG. See also the responses to Comments N3-2 and N3-12 for similar discussion pertaining to the effects and consideration of safety culture in this study.

Comment Number: N3-5
Comment: The expert panel was asked in the elicitation process to provide their best understanding about how the frequencies of flow rates from breaks would change over time and circumstances instead of being asked their understanding of pipe break frequencies as a function of break size. The panel did not have any thermal-hydraulic experts on it. It appears that at times panelists were confused or unaware of the implications of their choices for ratios that reflected their understanding of how flow rates would change. For example, the report says quantitative comparisons had to be provided among piping systems and non-piping components so that the panelists could understand the implications of their responses. Often panelists did not know or calculate the total LOCA frequencies associated with their choices to understand the interdependencies of their answers.

Response: The LOCA Categories 1 to 6 were defined by the flow rate emanating from the break to be consistent with historical definitions for small, medium, and large break LOCAs (developed initially in NUREG-1150) which serve as the basis of many PRA models. These historical definitions are aligned with emergency core cooling system (ECCS) operation to mitigate each size of break. While the mitigation response is plant-specific, the flow rates corresponding to each LOCA category are generally representative.

Because the elicitation panel did not contain any thermal-hydraulic experts, the panelists were instructed to utilize the correlations between the flow rate and break size for BWR steam and liquid lines, and PWR liquid lines provided in Table 3.8 of NUREG-1829. The panelists simply had to substitute the break size in Table 3.8 for the LOCA category to provide their elicitation responses for any LOCA sensitive piping or non-piping system. All panelists appeared to clearly understand this requirement (i.e., there were no questions or confusion expressed before, during, or after the elicitation), and evaluation of their responses indicated that the correlations were correctly applied.

Comment Number: N3-6
Comment: It is not clear why the elicitation used different correlations (e.g., flow rate versus break size) than previous elicitations.

Response: The correlations between break size and flow rate used in this study are different from those used in NUREG/CR-5750 and in other past LOCA efforts. The earlier correlations date to the NUREG-1150 plant risk study. The BWR correlations used in NUREG-1150 were developed from a matrix of calculations conducted using existing thermal-hydraulic codes specifically developed for each plant studied in NUREG-1150. A variety of break areas were postulated in certain systems and different combinations of mitigation equipment were assumed to be operational. Calculations were performed for each plant type. The reported correlation of break area to flow rate is an amalgam of these results. It is assumed that the PWR correlations are similarly based, although their development is not as well-documented.

As discussed in Section 3.7 of NUREG-1829, these prior correlations were not adopted for this study because there was concern about their generic applicability given the plant specific nature of the calculations performed in NUREG-1150. There is also concern about the accuracy of the thermal-hydraulic codes existing at the time of the NUREG-1150 study. Discharge leak rates are highly uncertain. They are difficult to estimate and are a function of upstream conditions, break location, plant configuration, and mitigation reliability, as well as break area. Based on these considerations, the simple closed-form approximate solutions are sufficient for the generic correlations required for this elicitation. Application of the elicitation LOCA frequency results may require plant specific calculations to evaluate break flow rate history and required mitigation response for assumed break locations and sizes.

Comment Number: N3-7
Comment: The report is confusing in that in places it says rare events such as large seismic events, severe water hammer and other loading sources are not considered in the elicitation, while in others it displays results for LOCAs due to seismic events. Similarly, the report in one place says panelists were told "to not answer questions in areas where they had little or no expertise." Yet in areas where panelists did not provide input that was needed, it is stated that the NRC staff filled in the missing data based on the panelist's overall answers.

Response: There are indeed a number of places where the report displays results for LOCAs due to seismic events. Those results, however, are presented more for ancillary or supplementary purposes than as results that impact the findings (i.e., LOCA frequencies) from this effort. For example, seismic-induced LOCA frequencies are presented in Section 1.2 (Previous Approaches for Estimating LOCA Frequencies) and Section 4.3 (Piping Base Case Sensitivity Studies). The results in Section 1.2 are presented to provide historical perspective more than anything else. The base case sensitivity studies in Section 4.3 were conducted to examine the effect of a number of factors. In addition to evaluating the effect of seismic loads on selected base case frequencies, the effects of in-service inspections (ISI) and leak detection strategies, weld overlay repair (WOR), IGSCC mitigation, applied loading magnitudes, hydrostatic testing, and degraded material properties were also studied. Appendices D and F also contain more detail about the effect of severe or extreme loading (i.e., seismic) on the base case frequencies.

The purpose of these sensitivity studies was to document the impact of variables that affect LOCA frequencies in order to help the elicitation panelists develop their responses. The results from the base case analyses, which were used by the individual panel members to anchor their elicitation responses, did not include the effect of seismic loading. Section 3.5.1 of NUREG-1829 clearly states that the loading spectrum for each of the base case systems was assumed to be the normal steady-state and transient loading histories expected over 60 years of operations. Therefore, the seismic studies have no direct bearing on the overall findings and conclusions from the elicitation.

The second point made in the comment implies that data was imputed by the facilitation team to fill in for expertise gaps in panelist's responses. This is not true. Imputed data was used occasionally to fill in gaps in the elicitation responses, but these gaps were generally places where the panelist had not provided complete coverage intervals with every elicitation response. In this case, as described in Section 3.10, additional assumptions were sometimes necessary to fill in missing information in the panelists' responses. The facilitation team made assumptions or added missing information, whenever possible, in a manner consistent with the philosophy, approach, assumptions, qualitative insights, and other quantitative results already provided by the panelist. Whenever it was necessary to develop imputed data, this data was subject to panelist review and acceptance prior to final use." Imputed data was never used to compensate for individual panelist's gaps in expertise.

Key elements of this response have been incorporated into Sections 3.5.1 and 4.3 of the revised NUREG.

Comment Number: N3-8
Comment: The conclusion trainers wished to convey to the panelists during the training period is that a group can guess the answer to general almanac type questions better than they can individually does not directly apply to scientific questions. That is, having a group of smart individuals together and asking them for their opinion of an area where they do not directly know the answer does not mean that their summed answers will be closer to reality than those of any particular individual involved or that the group somehow will have better knowledge than the members have individually when the issue of concern is unknown science.

Response: The premise on which the elicitation is based is the empirical observation that a group estimate of a quantity about which its individual members have only vague information can be surprisingly close to the "truth". (Here, a group estimate means some value near the group center.) This empirical observation is documented in a 2004 book, *The Wisdom of Crowds* (Doubleday) by James Surowiecki, which reports on many examples of this phenomenon. One example involved the disappearance of the U.S. submarine *Scorpion* in 1968 in which a diverse group of technical experts (e.g., submarine experts, mathematicians, salvage divers) independently estimated the *Scorpion*'s location by extrapolating from its last known speed and position. The individual estimates were combined into a group estimate, which was only 220 yards (200 meters) from the *Scorpion*'s location when it was subsequently found.

Part of the elicitation training of the panelists consisted of an exercise in which the panelists were asked "almanac-type" questions about chronic disease statistics (see Section 3.3.2 and Appendix C). To maximize their relevance to the LOCA frequency questionnaire which each panelist subsequently completed, the training exercise questions had a similar structure to the LOCA frequency questions. The training exercise responses were then combined into group estimates. The results of the training exercise supported the empirical observation discussed in above book.

In both the *Scorpion* example and the almanac training, the technical experts were asked to provide estimates only loosely related to their areas of expertise. These situations are directly analogous to the LOCA frequency elicitation since these estimates were also not directly applicable to the panelists' expertise in three important ways. First, virtually all of the questions ask the panelists to extrapolate well beyond their knowledge and experience with passive system failure. Secondly, the questions are all concerned with determining absolute frequencies of passive system failure, a very different concept than failure mechanisms themselves. Finally, the questions all ask the panelists to estimate uncertainties associated with these frequencies, a concept which is even further removed from the phenomena. Therefore, the elicitation questions ask the panelists to estimate quantities about which they have only vague information, analogous to responding to almanac-type questions (and the *Scorpion* example) which has been used, in part, to validate the elicitation process

Additional discussion on the use and accuracy of elicitation procedures has been added to Sections 1 and 5 to clarify the issue raised by this comment.

Comment Number: N3-9
Comment: Base cases were developed by a base case team drawn from among the panelists. All other LOCA estimates are less certain than the base case, since all other numbers are derived from ratios to the base case median, 5th, and 95th percentiles. Base cases are not anchored in "real" LOCA frequencies, but are an extrapolation based on engineering judgment.

Response: The base cases do represent an extrapolation of the operating experience based on engineering judgment. This is not only appropriate, but necessary, give the paucity of operating experience. Leaking and through-wall cracks frequencies provide the most direct comparison between predictions and the operating experience. One of the base case analysts, Bengt Lydell, determined the through-wall crack frequencies for the base case systems based on the operating experience. These are compared with the PFM predictions conducted by David Harris. See Section 4.1 of NUREG-1829 for more details.

All other LOCA frequency estimates that were derived by anchoring relative responses to the base case scenarios do necessarily encompass greater uncertainty. However, this increased uncertainty was naturally captured by combining uncertainty distributions to calculate the final estimates. See Section 5 of NUREG-1829 for more details.

Comment Number: N3-10
Comment: There was no peer review of the elicitation and its results. It was decided that such a review would take too much time and resources. The process of performing the elicitation was reviewed by two external reviewers, who did not evaluate the questions asked of the panel by the facilitation team. The external review did not evaluate the accuracy or adequacy of the technical analyses performed to come up with LOCA frequencies.

Response: There was external peer review of the elicitation and the results. Two external reviewers reviewed the approach used to calculate the group estimates, and specifically addressed the accuracy and adequacy of the technical analysis performed to estimate the LOCA frequencies. This review included the analysis used to calculate individual panelist LOCA frequency estimates from each elicitation responses, as well as the methods used to combine individual responses into group estimates. One of the external peer reviewers, Dr. Alan Brothers, also has extensive experience with conducting elicitations and summarized his review on the framework of the elicitation as follows:

> "The process used to elicit frequencies from multiple experts and combine them into overall estimates was carried out in a manner consistent with accepted engineering practice. In particular, the methodology used to obtain frequencies from individual experts was of sound design and execution. Notable aspects of the methodology are the selection and training of experts, the identification of issues, and finding meaningful response modes for small probabilities; all of which lend confidence to the validity of the process. The iterative process, in which the experts were given opportunities to interact with each other and to understand the underlying reasons for differences of opinions, and given the opportunity to revise their opinion afterwards, also lends validity to the process. Training on almanac questions to help avoid overconfidence was also a positive step in the process. The authors show evidence of a good working knowledge of the literature on the psychology of probability elicitation, and the potential pitfalls of well known

> heuristics and biases. The variables were well defined and broken out into a sufficient level of detail for which meaningful judgments could be given. The use of ratios relative to a base case overcomes the problems people have in making small probability estimates. In sum, the process used to obtain frequencies from individual experts was sound in both the overall approach and in the details."

More detail is provided in the complete review as documented in ML051430431.

Comment Number: N3-11
Comment: The report assumed that since "there is little supporting data for the [panel] members to assess the absolute frequencies of LOCAs ... assessing relative frequencies should lead to more accurate results." This assumption may be false.

Response: An assessment of the relative likelihoods for comparable LOCA frequency contributors is a more natural reflection of knowledge and experience than is an assessment of the absolute frequencies of each contributor. Relative comparisons require less extrapolation from the experience and knowledge base of the panelists, and are appropriate when estimating the frequency of rare events. This approach was explicitly endorsed by external reviewer Dr. Alan Brothers (Reference ML051430431), who is knowledgeable about the decision analysis literature and best elicitation practices. The accuracy of using relative rather than absolute frequencies is discussed in Sections 3.3 and 3.8.1 of the NUREG, and is supported by the results of an elicitation training exercise conducted with the elicitation panelists. See also the response to Comment N3-3 for additional discussion.

Comment Number: N3-12
Comment: On pages 3-9 and 3-10, the issue of a plant's safety culture and the potential impact on LOCA frequencies is considered. The report says that safety culture is cyclic and can significantly impact LOCA frequencies, but then dismisses the impact because it is "generic." If safety culture can "significantly impact" LOCA frequencies, it would seem that this factor should either be included in the estimates or, at least, in the uncertainty bounds. Also, it is not clear how one should go about assessing the impact from safety culture when developing LOCA frequency estimates for a specific nuclear plant.

Response: The safety culture question was structured so that each panelist was required to address relative changes in the future safety culture resulting from issues raised during the elicitation. Possible future changes were assessed relative to the current existing safety culture. The panel members generally expressed the belief that the future median safety culture will not differ dramatically from the current culture. In fact, most participants expected a small improvement due mainly to continued experience and continued technology advances.

However, as discussed in Section 6.2 of NUREG-1829, many panelists expressed the opinion that safety culture is a highly plant-specific factor. While most panelists believe that the industry is generally acting in a consistent, safety conscious manner, individual plants can deviate markedly from general industry philosophies and practices. The vessel head degradation at Davis-Besse was a commonly cited example. Panelists argued that the LOCA frequencies at these less safety conscious plants could be greatly elevated compared with the remaining population. However, this trend would not affect either the median LOCA frequencies or safety culture estimates, but would more likely affect the upper bounds of these distributions.

The upper bound LOCA frequency estimates do incorporate broad differences among plants related to important variables (i.e., plant system, material, geometry, degradation mechanism, loading, mitigation/maintenance) in determining both the generic LOCA frequency estimates and especially the uncertainty bounds (see responses to Comments N1-3 and N2-1). The upper bounds do not, however, account for egregious safety culture deficiencies that may exist at any single plant. It is understood that unique plant features and safety culture may also influence LOCA frequencies and plant-specific estimates may fall outside the generic fleet estimates. For instance, several panelists indicated that a deficient safety culture at any specific plant could, if unchecked, result in LOCA frequency estimates that are 1 to 2 orders of magnitude higher than the median estimates. However, the elicitation panelists generally expect that accident frequency increases stemming from deficient safety practices will continue to be identified and mitigated through regulatory oversight policies and procedures. The possible increases in LOCA frequencies stemming from deficient safety culture therefore serve as more of a warning for the agency to remain vigilant in addressing these issues rather than a prediction.

Because there was a general expectation among the panelists that future safety culture changes would not affect the median LOCA frequencies, and that accounting for possible deficiencies at single plants would have skewed the distributions such that they would not represent generic values, the LOCA frequency estimates were not adjusted to account for safety culture effects. This decision was endorsed by the elicitation panelists.

See Section 7.1 of NUREG-1829 for additional discussion on safety culture effects. See also responses to Comments N3-2 and N3-4 for similar discussion pertaining to the effects and consideration of safety culture in this study. Key elements of this response have been incorporated into the Executive Summary and Sections 2, 6.2, 7.1, and 9 of the revised NUREG.

Comment Number: N3-13
Comment: The NUREG provides details of bases including the methods, assumptions, and results for the individual estimates of only four of the panelists who provide base case results. A short summary of the general approach to making estimates is provided for the other panelists, except for G. DeBoo. This information for the remaining eight panelists does not form an adequate basis for concluding that the estimates were made with necessary rigor. The details of the methods, assumptions, and results for these eight panelists should be provided.

Response: G. DeBoo never provided a quantitative response to the elicitation. He participated in the elicitation and provided some qualitative insights but did not provide any quantitative results. This is why his general approach is not included in Appendix K. Appendix K provides the general philosophy and approach followed by the other 11 panelists for their individual elicitations. The expanded discussion provided in Section 3.5.1.2 for the four panelists who provided estimates for the base cases describes how these estimates were calculated. However, this discussion does not detail the approaches or philosophies that these panelists used in developing their individual elicitation responses. The intent of the descriptions in Appendix K was to provide the philosophical approach used by each of the panelists in arriving at their responses. The level of detail needed to support the analysis was determined by each panelist. Each panelist discussed their approach and the basis for their elicitation responses in their individual elicitations sessions.

Comment Number: N3-14
Comment: The NUREG defines the LOCA categories in terms of discharge flowrate. However, it should be clarified that the panelists predicted the frequencies of certain size breaches in the reactor coolant system, regardless of flow rate. Does the flow rate information consider that some completely broken pipes have discharge from both ends, or is it consistently calculated based on discharge from only one end?

Response: As discussed in the Section 3.7, the panel expertise was concentrated in the areas of structural analysis, materials, and fracture mechanics, not thermal-hydraulics. Therefore, it was more natural for the panelists to develop their responses as a function of the effective break area. In order to facilitate this evaluation, it was necessary to develop a correlation between the makeup flow rate and effective break area. The LOCA size categories were defined as cumulative frequencies at a given flow rate; these flow rates were then converted to an equivalent pipe diameter. The LOCA frequencies associated with each LOCA size category relates to the cumulative frequency of a single-ended break of the cited size, and all breaks (including double-ended breaks) of that size and larger.

Comment Number: N3-15
Comment: In Tables B.1.7/8, there are some pipe sizes which appear to be in error (e.g., the 24-inch reactor water cleanup and the 28-inch safety/relief valve discharge). Also a survey performed by the staff indicates that the maximum surge line size is 16 inches (not 14 inches), and the maximum PWR residual heat removal is 16 inches (not 12 inches). Additionally, should the piping sizes reflect the internal dimensions since those determine the LOCA categories?

Response: The data presented in Tables B.1.7 and B.1.8 provides information developed by the panelists at several brainstorming sessions. Therefore, they represent the panelists' recollection of the geometries rather than a systematic analysis of piping system geometries. Therefore, in some cases, there are minor discrepancies between the pipe size ranges provided and actual plant configurations. For example, there are 16-inch diameter surge lines, such as those at South Texas Units 1 and 2, while Table B.1.8 indicates that the range in surge line diameters is 10 to 14 inches. However, these relatively small discrepancies in the outer ranges of these bounds have no significant effect on the results and conclusions which focused on calculating generic LOCA frequency estimates. The diameters provided in this table do represent nominal pipe dimensions (e.g., 12-, 16-, 28-inch diameter) for the sake of simplicity, rather than inside pipe dimensions. This convention was utilized by consensus among the panelists.

Key elements of this response have been incorporated into Appendix B of the revised NUREG.

Comment Number: N3-16
Comment: Panel members were allowed to ignore the approach proposed by the facilitation team for coming up with ratios of how flow rates would vary for various cases and could choose their own methods.

Response: It is unclear whether this comment refers to the flow rate correlation or elicitation framework. Correlations were developed to relate the flow rate to the effective break diameter for BWR liquid and steam systems as well as PWR liquid systems. These correlations were fixed and panelists did not develop ancillary expressions. Conversely, the elicitation framework was structured to allow the panelists maximum flexibility in providing their quantitative responses within the overall structure as embodied in the questionnaire. In fact, no two panelists used exactly the same approach. Some chose to modify the base cases developed by the panel or even develop their own base case scenarios. A few panelists provided absolute values instead of the requested relative values (ratios) and these were converted to equivalent ratios for analysis. Regardless of the different paths taken by the individual panelists, the goal was to calculate comparable LOCA frequency estimates using an approach that promoted confidence in the responses provided.

Key elements of this response have been used to clarify Sections 3.7 and 3.9 of the revised NUREG.

Comments Related to Section 4 of NUREG-1829

Comment Number: N4-1
Comment: One set of base cases (B. Galyean) partitioned the estimated LOCA frequencies among categories "using the common assumption that the frequency of a LOCA decreases as pipe size increases.

A somewhat arbitrary scaling factor of ½ order of magnitude was used to scale successive LOCA categories by assuming a lognormal probability distribution on LOCA frequency." Assumptions are piled on assumptions. An alternative assumption could be that the frequency of LOCAs approaches some limit as pipe size increases rather than being monotonically decreasing. Similarly different assumptions about distributions could alter the base case. Some of the results for those who used PFM techniques were adjusted through manipulation of variable input parameters in order to more closely match the service history through-wall cracking frequencies. This raises the question of why should PFM be considered useful as a predictor for future LOCAs (some 30+ years from now) or for break sizes where there is no data, if the analysis method does not even do a good job of predicting existing service history results.

Response: The half-order of magnitude (i.e., approximately a factor of 3) decrease in frequency for each increase in LOCA size is an assumption based on the general practice employed in estimating LOCA frequencies over the past 30 years starting with the Reactor Safety Study (WASH-1400, 1975). This assumption is supported by work done by Beliczey and Schulz [Beliczey, S., and Schulz, H., "Comments on Probabilities of Leaks and Breaks of Safety-Related Piping in PWR Plants," *International Journal of Pressure Vessel and Piping*, Vol. 43, pp. 219 – 227, (1990).]. In this paper, a combination of operating experience and fracture mechanics is used to demonstrate the well-established insight that the conditional probability of a rupture, given a leak, decreases as pipe diameter increases. This insight is applicable because the size of detectable cracks and leaks remains relatively constant as a function of pipe size. Therefore, the relative crack or leak size as a function of the pipe circumference decreases, and the safety margin increases, as the pipe diameter increases. Therefore, there is no basis to expect that the LOCA frequency will approach a limit as pipe diameter increases as suggested by the commenter.

Additionally, Beliczey and Schulz developed a quantitative conditional failure probability --- which decreases by ½ order of magnitude for each successively larger LOCA size --- that was based on the propensity of through-wall fatigue flaws to lead to successively larger LOCA sizes. Although the quantitative conditional failure probability is not applicable to all failure mechanisms and systems, this simple relationship has been extensively employed. This assumption was also employed by Galyean in his analysis but the three other base case team members used separate conditional failure probability relationships or calculated them directly from PFM analysis. As indicated in Section 4.2, the other base case results support the use of Beliczey and Schulz' quantitative conditional failure probability as a general rule-of-thumb.

All assumptions associated with the base case calculations were clearly detailed to the entire panel so that the panelists had a clear understanding of the conditions, limitations, and appropriateness of the base case results. Panelists, in their elicitation responses, were required to determine their own conditional failure probability relationships among successively increasing LOCA sizes. Hence, they were not bound by assumptions or analyses performed during the base case development.

The commenter also questions the usefulness of PFM as a predictor for future LOCA frequencies. There is a fundamental difference between operating experience and PFM predictions. Operating experience does reflect actual passive system performance, but the conditions associated with each passive flaw (or degraded condition) that affect the propensity of a component to crack are generally not completely known. Therefore, determining the effect of important variables is difficult, if not intractable. Conversely, PFM predictions have been demonstrated to be accurate for comparison of experimental results where the model inputs are well characterized. These codes are particularly accurate in predicting the relative changes in failure rates that occur upon changing important variables. See Section 1.2.2 for additional description of differences in these approaches.

Because the variables associated with the operating experience are not well-known, direct comparison with PFM results for a specific set of variables is not possible. However, benchmarking the relative changes predicted by PFM codes using operating experience for leaks and small LOCA frequencies (where the operating experience is most meaningful) is a method which marries the strengths of both approaches. This has been performed and is described in Section 4.1 of the NUREG. This is also the technique that many of the elicitation panelists used. Often panelists' LOCA Category 1 estimates were based on operational experience while PFM was used to establish trends for extrapolating the Category 1 current-day estimates to both larger LOCA sizes and into the future.

Key elements of this response were incorporated into Section E.2 of Appendix E of the revised NUREG. See the responses to NGC4, NGC5, N3-1, and N4-11 plus Comment 4-1 from Appendix M of the revised NUREG for additional information related to the use of PFM within this elicitation.

Comment Number: N4-2
Comment: It is understood among engineers that it is much more reliable to interpolate than to extrapolate. At the NRC in dealing with issues such a reactor fuel burn-up, departure from nucleate boiling, and various other phenomena, we normally do not credit extrapolated analyses, and make licensees provide ample data over the range of interest if they want credit in licensing space. Panelists D. Harris extrapolated from through-wall pipe crack lengths to LOCAs for his base case estimates. V. Chapman in his base case estimates used expert judgment to extrapolate elastic crack opening displacement beyond the elastic limit and used expert judgment to assess the defect length distribution at failure.

Response: Expert elicitation is only used for quantifying phenomenological knowledge when data or modeling approaches are insufficient. Thus, expert elicitation is not a methodology which would normally be used in situations where interpolation is possible. Instead it is used in situations where one is forced to extrapolate beyond the existing data. When estimating frequencies associated with rare events, such as LOCAs, extrapolation is necessary because the data do not exist which lends itself to interpolation. The uncertainty bounds associated with each elicitation point reflect the increased error which results from an extrapolation scheme such as this. See the Executive Summary and Section 1.5 for more details.

Comment Number: N4-3
Comment: The report considers factors of 5 to 8 to be "relatively good agreement" between the service history and PFM for two base cases, but there was significant disagreement between the service history and PFM results for the other three base case results. Why should we believe these results somehow capture physical reality when they exhibit such large disparities? The explanations given in the report regarding these differences appear to be unclear. In one base case, "the PFM estimate is six orders of magnitude less than the service history estimate. Reasons for the difference are not readily apparent." Even with these disparities, the report goes on to say that the "base case estimates ... are those that the base case team members believe are most accurate." Regarding comparison between the PFM base case modelers for one case, the report notes positively that for this one case, "for these more similar conditions [the LOCA frequencies] **only** [my emphasis] vary by three orders of magnitude, instead of seven." It is not clear if similarity of results among cases or evaluators is more of a coincidence rather than any convergence of estimates. It is apparent that there is little understanding of where reality lies for these calculations. Who knows which, if any, of the calculations is correct or even how large the uncertainty is?

Response: The first part of the response addresses the commenter's concern about the lack of understanding of where reality lies for the base case calculations. According to the elicitation panelists, the base case results developed using operating experience to estimate the 25-year (i.e., current-day), Category 1 LOCA frequencies provide the most realistic starting point. As described in Section 6.1 of the NUREG, most panelists chose to anchor their elicitation responses to one or both of the operating-experience-based base case analyses (Sections 3.5 and 4) for making the current-day LOCA frequency estimates. For future estimates, a number of the panelists augmented the operating-experience-based estimates with results from the PFM studies and associated sensitivity analyses. However, that almost always involved making relative assessments with PFM models instead of using PFM analyses in an absolute sense.

The commenter also questioned, in light of the large discrepancies in the results, the suitability of the statement that these base case estimates "*… are those that the base case team members believe are most accurate."* The context of this statement (Section 4.2) is that the base case team developed a number of base case analyses over time. As the base case definitions and the estimates evolved over time, the base case team members came to believe that their latter estimates (those reported in Table 4.1) best reflected the base case definitions. This assessment by the base case team members was necessary because the Table 4.1 base case values were not available to the panel members during the initial individual elicitations. At that time, the panel members only had interim base case frequencies. However, panel members modified their original elicitation responses once the final base case results (as reported in Table 4.1) were completed and made available. The NUREG has been revised to clarify this statement and the surrounding context.

Finally, the statement quoted by the commenter about the large differences in the PFM and service history estimates and the lack of understanding for the discrepancies was taken from the draft NUREG. This portion of the NUREG (Section 4.2) has been significantly revised and the differences among the estimates are accurately noted for each base case. Furthermore, reasons for the discrepancies are provided. This rationale was not included in the draft NUREG because resolving the discrepancies was not an objective of the elicitation. Rather, the discrepancies were presented to the panelists to inform their base case selection. However, based on this comment, the NUREG was revised to inform the reader of the causes for these discrepancies.

Also see the responses to Comments GC5 and GC5a for additional discussion on the use of PFM codes by the panelists in their elicitation responses.

Comment Number: N4-4
Comment: The Section 4.4.2 discussion of pressurized thermal shock (PTS) base case disregarded the important effect that small LOCAs could become large LOCAs due to the injection of cold water at high pressure into the downcomers causing PTS.

Response: As described in Section 3.5.2.3 of NUREG-1829, the PTS challenges are appropriate to consider in the expert elicitation because they could lead to vessel failure and a resulting large LOCA. However, some care is required in this assessment. The biggest PTS risk contribution is usually from SB, MB, and LB LOCA initiating events, especially as material embrittlement increases. Further, the PTS risk is proportional to the LOCA frequencies. Because of this dependency, the total PTS risk cannot be used as a base case for determining the frequency of other pressure vessel or non-piping failures. However, there are other loading transients which contribute to PTS risk. These transients include stuck open primary and secondary valves and feed and bleed operations. The stuck-open valve and feed and bleed transient contribution to PTS risk ranges from 1% to 80% depending on the plant type and the embrittlement level. The panelists were only asked to consider the frequency associated with PTS risk from these transients for possible anchoring during the elicitation process.

Therefore, although it is recognized that significant PTS risk results from SB, MB, and LB LOCA initiating events, only the risk associated with non-pipe break failures was quantified to ensure that these base case estimates are not dependent on the frequencies being estimated in the elicitation. Section 4.4.2 of the revised NUREG has been clarified to indicate that the base case results do not represent the total PTS risk, but only the risk from events other than passive system failures.

Comment Number: N4-5
Comment: It is unclear if boric acid wastage was considered for non-piping failures.

Response: Boric acid wastage was considered as one of the failure scenarios for non-piping systems. As discussed in Appendix B, the pressurizer shell failure scenario (see Table B.1.13) would most likely occur by boric acid wastage from the outer diameter of the shell. Similarly, the RPV failure modes include RPV wastage due to boric acid corrosion along with vessel head bolt failure, failure of CRDM connections, nozzle failure, and RPV corrosion fatigue. At least one panelist considered boric acid wastage to be a significant contributor to the overall LOCA frequencies. This panelist saw boric acid corrosion of the RPV head as a significant contributor (~40 to 50%) to the overall LOCA frequencies for PWR non-piping systems for Category 3 and 4 LOCAs.

Comment Number: N4-6
Comment: Given the huge uncertainties, Table D.1 shows too many significant digits.

Response: It appears that the reviewer is referring to an earlier version of the NUREG where sections were designated by letters instead of numbers. Table D.1 is actually Table 4.1 in final NUREG. The number of significant digits used for the results in the main body of the report is appropriate in light of the uncertainties associated with these results. However, the authors realize that there are an unjustifiable number of significant digits in a number of tables contained within the appendices. However, these appendices were contributed separately by the panelists and represent the results of their analysis. Because the authors' policy was to not alter the technical content of a panelist's appendix, the number of significant digits in tables within the appendices was not modified.

Comment Number: N4-7
Comment: The conclusion in Section 4.2 on Piping Base Case LOCA Frequencies that the average conditional probabilities between successive LOCA categories is approximately 0.5 is a circular argument. It was already stated in Section 3.5.1.2.2, B. Galyean's Base Case Estimates, that "a somewhat arbitrary scaling factor of ½ orders of magnitude was used [by Galyean] to scale successive LOCA categories by assuming a lognormal probability distribution on LOCA frequency." It is not surprising that one gets what one assumes.

Response: It is correct that one set of base case estimates developed using operating experience assumed conditional failure probabilities between successive LOCA sizes. Bill Galyean assumed that the base case frequencies decrease by a ½ order of magnitude for each successively larger LOCA category. This assumption was based, as described in the response to Comment N4-1, on precedent and supported by research by Beliczey and Schulz [Beliczey, S., and Schulz, H., "Comments on Probabilities of Leaks and Breaks of Safety-Related Piping in PWR Plants," *International Journal of Pressure Vessel and Piping*, Vol. 43, pp. 219 – 227, (1990).]. See Appendix E for more details.

However, the other operating experience estimates, provided by Bengt Lydell, developed conditional failure probability (CFP) relationships based on an analysis of operating experience trends. Therefore, these CFPs were not assumed a priori. While these conditional probabilities vary somewhat as a function of degradation mechanism and LOCA size, the average failure frequency decrease is approximately ½ order of magnitude between successive LOCA categories. See Appendix D for more details.

Additionally, the two PFM estimates (Appendices F and G) make no a priori assumptions about this CFP. The CFP is determined directly from the analysis. The fact that, on average, the PFM and the other operating experience CFP estimates reasonably approximate a ½ order of magnitude decrease between successive LOCA categories lends credence to the CFP assumptions made by Bill Galyean.

The base case estimates were provided to all members of the panel as part of a larger package of available background information. All assumptions associated with the base case calculations were clearly detailed to the entire panel so that the panelists had a clear understanding of the conditions, limitations, and appropriateness of the base case results. Panelists, in their elicitation responses, were required to determine their own conditional failure probability CFP relationships. Hence, they were not bound by assumptions or analyses performed during the base case development. Panelists that used any of the base case CFP estimates either implicitly or explicitly agreed with that portion of the analysis.

Section 4.2 of the revised NUREG has been clarified based on this comment. Also, see the response to Comments N4-1 and NE-1 for additional information.

Comment Number: N4-8
Comment: Section 4.3.6 Effect of Seismic Loading and Hydro Test does not provide generic insights, which would depend on the seismic hazard at the site, the codes to which the plant was built, which seismic response spectra it used, and perhaps even issues such as what damping factors were assumed.

Response: As described in Section 4.3.6, the impact of seismic stress was not explicitly considered within the elicitation because the principal objective of the elicitation (Section 2) was to estimate LOCA frequencies under normal operating loads and expected transients. The authors agree with the commenter that Section 4.3.6 does not provide generic seismic insights. However, the objective of this sensitivity study was not to provide generic seismic insights, but to determine the effect of an assumed seismic load history on these base case results. The seismic magnitude for this sensitivity analysis was selected from the design seismic load spectrum developed for a specific plant. In this particular case, the piping system considered was the hot leg-pressurizer weld subjected to a seismic loading up to a 5SSE level. The stresses are discussed in Appendix F, Section F.3.2.2, which provides the original reference for these stresses (Original PRAISE development NUREG/CR-2189, Vol. 5). However, for the hot leg, there is little effect from the assumed seismic loading due to the small SSE stresses that are applicable for the hot leg configuration modeled. These stresses are plant specific and other configurations, resulting in higher stresses, could have a greater impact on the LOCA frequencies.

Key elements of this response were incorporated into Section 4.3.6 of the revised NUREG. Also see the responses to Comments NGC10 and N3-7 for additional discussion about the seismic sensitivity analyses.

Comment Number: N4-9
Comment: The frequency estimates in the sensitivity studies shown in Section 4.3.7 are not believable. The effect of fracture toughness on cumulative LOCA probabilities (Note: it probably should read, "frequencies") for the PWR-1, (i.e., smallest LOCAs), Figure 4.16, has expected frequencies tens of thousands of times older than the universe.

Response: The results for the sensitivity study on the effect of fracture toughness on the LOCA probabilities are indeed extremely low. The reason is that the only degradation mechanism considered was thermal fatigue from pre-existing defects. The only cyclic stresses considered in this fatigue analysis were 3 heat-up and cool-down cycles each year (with a 6.5 ksi [45 MPa] stress range) plus the seismic stresses which were generally quite low, i.e., 1.96 ksi (13.5 MPa) for an SSE event if the seismic event contained 200 stress cycles all of the same magnitude.

These stresses result in very little crack growth over the operating history which leads to the low absolute frequencies shown in Figure 4.16. The implication is that failure is unlikely due to this specific mechanism under the assumptions and conditions modeled in the analysis. The point of the figure is simply to show the relative changes in the results due to changes in material fracture toughness for this one set of analysis conditions. While the absolute frequencies are not particularly meaningful (other than the fact that they are incredibly low), the relative differences caused by different fracture toughness values is illustrative.

Key elements of this response were incorporated into Section 4.3.7 of the revised NUREG. See also the response to Comment NGC10 about the interpretation of extremely low frequencies from the PFM and other analyses.

Comment Number: N4-10
Comment: On page 4-1, the NUREG states that, "... base case frequencies were developed by a subset of the panel. These frequencies were then provided to the other panelists as possible anchoring frequencies for use during their elicitations. Panelists had the option of using the base cases and associated frequencies or developing an alternative approach." From the statistical point of view, the elicitation process is essentially the selection of a random sample. A random sample, in statistics, is chosen in the context of an experimental design so as to achieve favorable inferential properties, such as for example minimum variance.

The NUREG should explain what the "experimental design" described by the reported process is, and what are the inferential properties of the sample associated with this design. For example the discussions in Section 1.5 of the report do not explain why this particular elicitation yields an unbiased estimator of the median LOCA frequency and the attendant 95th percentile.

Response: It is not true that, as stipulated in this comment, "the elicitation process is essentially the selection of a random sample". As discussed in Sections 1 – 3, the elicitation process is carefully structured to calculate LOCA frequency estimates, supported by qualitative rationale, using expert opinion. The accuracy and appropriateness of using an elicitation process to estimate quantities, such as LOCA frequencies, where the data is sparse and sufficiently accurate models do not exist is well established (See Section 1). The elicitation panel was also chosen to be broadly representative of the technical community, rather than a random sample from this community. Additionally, as stated in Section 3.3 of the NUREG, "a basic premise in using an expert elicitation process is that the panel responses as a whole have no significant systematic bias". See Section 3.3 and Section C.3 of Appendix C for a discussion of this issue. Also, see the response to Comment N5-17 for additional discussion about the assumption of no systematic bias in the elicitation responses.

Comment Number: N4-11
Comment: On page 4-1, the NUREG states: "The probabilistic fracture mechanics (PFM) results are quite sensitive to initial input variable distributions for these base cases. Relatively minor changes can result in through-wall cracking frequency differences of several orders of magnitude (Appendix F)." It goes on to say (page 4-2): "The PFM estimate is approximately six orders of magnitude less than the service history estimate [for the feedwater system base case, BWR-2]." The discussion notes "substantial" variability between two PFM-based studies. Taken together, these statements seem to say the PFM models are not useful for estimating absolute LOCA frequencies. However, PFM methods were used by some panelists in developing their input to this study.

Response: The authors largely agree that PFM models, by themselves, are not accurate in estimating <u>absolute</u> LOCA frequencies. This is one reason why the elicitation was conducted (See Section 1). However, PFM methods were not solely used by any panelists in developing their elicitation responses as stated in the comment. The PFM analyses were generally used to provide relative comparisons to support trending analyses while operating experience estimates were used to anchor the base case results. See responses to Comments NGC4, NGC5, N3-1, and N4-1 and also Comment 4-1 from Appendix M of the revised NUREG for additional information regarding the use of PFM in the elicitation.

Key elements of this response were incorporated into Section 4.2 of the revised NUREG.

Comment Number: N4-12
Comment: In several places of the NUREG (e.g., Table 4.1), it is indicated that PWR LOCA Category 5 includes the pressurizer surge line piping. However, the Category 5 equivalent break size opening is 14 inches and greater, while the pressurizer surge line piping inside diameter ranges from about 215 mm (8.5 inches) to 325 mm (12.8 inches). Therefore, it appears that breaks in the pressurizer surge lines should be included in LOCA Category 4 and smaller. If pressurizer surge line breaks were considered as double-ended break areas, this should be explained.

Response: The break size versus flow rate correlations shown in Table 3.8 for BWR: Steam, BWR: Liquid, and PWR: Liquid are based on the assumption of a single-ended break at the cited size. Therefore, the break sizes in Table 3.8 correspond to a partial fracture for pipes with larger diameters than the break size, a complete single-ended rupture in pipes with the same inside diameter, or a DEGB in pipes having inside diameters $1/\sqrt{2}$ times the break size. The effective break size necessary for a Category 5 LOCA, from Table 3.8 in NUREG-1829 is 14 inches. Most surge lines have 10 to 14 inch nominal outside diameters. However, the largest surge line (South Texas) has a 16 inch outside diameter and a 12.8 inch (325 mm) inside diameter. Therefore, as indicated in the comment, a complete single-ended rupture would not cause a Category 5 LOCA.

However, if a double-ended rupture occurred in the pressurizer surge line, this would effectively separate the initial pressurizer inventory from the rest of the RCS. If the inside pipe diameter is 10 inches (250 mm) or greater (which encompasses most surge lines), this rupture would qualify as a Category 5 LOCA (Table 3.8) until the pressurizer itself completely drains. The pressurizer inventory should be depleted within several seconds after the break. For the remainder of this hypothetical transient, RCS inventory loss would occur only through the hot-leg side of the break, and the flow rate would be representative of a LOCA Category 4 break. In summary, a DEGB of the surge line that is greater than 10 inches (250 mm) would initially be a Category 5 LOCA for several seconds and then would decrease to a Category 4 LOCA after the pressurizer has drained.

Comments Related to Section 5 of NUREG-1829

Comment Number: N5-1
Comment: All distribution shapes were assumed by the facilitation team to be lognormal, which is the most convenient for statistical manipulation.

Response: The motivation for assuming a lognormal structure for the panelists' responses was the structure of the elicitation questions. Because all of the questions called for a ratio as a response, the lognormal distribution was a natural choice. An external reviewer of the statistical methodology also concluded that the use of the split lognormal distribution was a reasonable and appropriate assumption. While many sensitivity analyses were performed, none were performed specifically to evaluate this assumption. Possible alternative distributional forms are the loguniform or the logtriangular. However, variations in the individual LOCA frequency estimates resulting from replacing the lognormal with other plausible distributional forms are expected to be significantly smaller than the differences between the individual estimates. Accordingly, because of the very large diversity in the individual estimates, it is not expected that the particular choice of the distributional form leads to group estimates that are significantly different than those based on the lognormal assumption.

The rationale for choosing the lognormal structure has been added to the revised NUREG in Section 5.3.1.

Comment Number: N5-2
Comment: There was no effective discussion or justification of the basis for assuming the panelists' answers were from a lognormal distribution.

Response: See response to N5-1.

Comment Number: N5-3
Comment: The statistical and other manipulations of the panelists' answers consistently sought to minimize or smooth over the effects of outlier results. For example, the use of the geometric mean for combining the answers of the panelists together damps down differences. Arithmetic means were used in NUREG-1150. Sentences in the report such as "the baseline approach was chosen to develop consensus-type results from the panelists' opinions in a manner consistent with the elicitation objectives. [emphasis mine]" can be seen as implying that divergent results are less credible. It appears as if when a panelist's results were at variance with those of others, such results were systematically dampened or minimized. It is clear that the goal was for answers to converge.

Response: As discussed in Section 5.5 of the NUREG, a fundamental assumption underlying the use of expert elicitation is that the elicitation responses are not systematically biased. It follows that the group opinion should be somewhere in the middle of the group, especially if there are wide differences between individual opinions. This conclusion has been validated by many exercises using almanac-type questions, including the training exercise reported on in Section 3.3 of the NUREG. For the specific numerical results in the NUREG, the geometric mean falls somewhere in the middle of the group while the arithmetic mean does not.

Contrary to the comment, outlier results were not "systematically dampened or minimized". Outliers contributed to the group estimates both through the geometric mean and through the confidence intervals for the group estimates.

Comment Number: N5-4
Comment: The facilitation team chose to use the geometric mean as a way to aggregate panelist responses, and was antagonistic about the use of the arithmetic mean, which was used in NUREG-1150 and in NUREG/CR-5750 for aggregation of individual responses. The elicitation report inappropriately downplays the reasons to use the arithmetic mean in an apparent desire to minimize the effect of "outlier' results.

Response[1]**:** The geometric mean rather than the arithmetic mean was used as the primary method to aggregate the individual estimates to calculate the group estimates because a principal objective of this elicitation study was to determine consensus-type estimates that approximate the median of the individual estimates (Section 2). A sensitivity study (Section 7.6.4.1) demonstrated that there is little significant difference between the median and the geometric-mean aggregated group estimates.

There are also precedents for using the median when the individual estimates differ by several orders of magnitude, as they do in NUREG-1829. The references listed below on using the median to aggregate expert judgments are incorporated in the final NUREG.

---

[1] It is important to note that NUREG/CR-5750 did not aggregate data or expert opinion to determine LOCA frequencies. LOCA frequencies were estimated solely from operational experience (See Section 1).

1. Vo, Truong V., Heasler, Patrick G., Doctor, Steven R., Simenon, Frederic A., and Gore, Bryan F., "Estimates of Rupture Probabilities for Nuclear Power Plant Components: Expert Judgment Elicitation," *Nuclear Technology*, Vol. 96, pp. 259-271, 1991.

2. Vo, Truong V., Simenon, Frederic A., Gore, Bryan F., and Livingston, James V., "Expert Judgment Elicitation on Component Rupture Probabilities for Five PWR Systems," *Reliability and Risk in Pressure Vessels and Piping*, PVP-Vol. 251, pp. 127-140, American Society of Mechanical Engineers, 1993.

3. "A Pilot Application of Risk-Informed Methods to Establish Inservice Inspection Priorities for Nuclear Components at Surry Unit 1 Nuclear Power Station ," NUREG/CR-1681, Rev.1, PNNL-9020, Rev.1, U.S. Nuclear Regulatory Commission, February 1997.

4. "Technical Elements of Risk-Informed Inservice Inspection Programs for Piping," Draft Report, NUREG-1661, U.S. Nuclear Regulatory Commission, January 1999.

While NUREG-1150 provides a precedent for using the arithmetic mean, the arithmetic mean is typically used to display the central tendency of a population distribution. An arithmetic mean of values that spans multiple orders of magnitude essentially discounts the lower half of the individual estimates and may be determined by the highest one or two values. Hence, arithmetic-mean aggregation does not approximate the median of the individual estimates and results in higher group estimates than geometric-mean aggregation. However, because alternative aggregation methods can lead to significantly different results than those obtained using the geometric mean, a particular set of LOCA frequency estimates is not generically recommended in the NUREG for all risk-informed applications. The purposes and context of the application must be considered when determining the appropriateness of any set of elicitation results. In particular, during the selection of the BWR and PWR TBSs for the proposed 10CFR50.46a rule making, the NRC staff considered the totality of the results from the sensitivity studies rather than only estimates calculated using geometric-mean aggregation.

Also see the response to Comment N5-3 and Section 5.6.4.1 of the NUREG for related information. The Executive Summary and Section 5 of the NUREG have been modified as a result of this comment.

Comment Number: N5-5
Comment: The report seeks "the group opinion" and states that the "middle of the group" is the correct place to choose the group opinion rather that using the arithmetic mean, which can be dominated by those values that are largest. Unfortunately, use of the geometric mean tends to hide the diversity of opinion or degree of uncertainty in results. The Commission continues to need a process that would allow it to make decisions using risk information other than a simple, single "best estimate" value. It is not clear how the uncertainty of information or diversity of opinion are to be included specifically in the decision-making process.

Response: Contrary to the comment, it is not true that "the use of the geometric mean tends to hide the diversity of opinion or degree of uncertainty in results". The geometric mean is used to provide group estimates of the bottom-line parameters (mean, median, 5^{th} and 95^{th} percentiles). As discussed in Section 7.5, the uncertainties are captured by the 5^{th} and 95^{th} percentiles and the panel diversity by the confidence bounds on the bottom-line parameters.

It is important to separate the estimates of the LOCA frequencies as provided by the NUREG from the application of these results to regulatory decision-making. As noted in the Executive Summary, the users of the study results are in the best position to apply the results to their particular applications.

Accordingly, it is beyond the scope of the study to consider how the results can be applied to regulatory decision-making.

Also see the response to Comment N5-4 for related discussion.

Comment Number: N5-6
Comment: While structured decomposition of a problem may result in risk estimates being more systematic and completely thought out, it may still be highly inaccurate.

Response: The expert panel devoted extensive effort to technical issue formulation and problem decomposition. This effort is described in Section 3.4 and the meeting minutes (see Appendix B) and is embodied in the elicitation questionnaire (see Appendix J). The panel consensus was that the questionnaire represents an accurate decomposition of the problem. It is not possible to gauge the relative accuracy of the estimates for large breaks given the lack of representative operating experience. However, the results correspond well with SB LOCA frequency estimates based on operating experience. Additionally, comparisons with historical LOCA frequency estimates are used to demonstrate how the quantitative responses and supporting qualitative rationale has resulted in changes that are justifiable based on the panelists' opinions. See Section 7.9 for this comparison.

Comment Number: N5-7
Comment: Many of the statistical manipulations made by the facilitation team on the answers made by the panel assumed that the answers were statistically independent among the panelists. There was no evaluation of whether there was a loss of statistical independence during brainstorming sessions and other group sessions.

Response: The assumption of statistical independence between the panelists used in the analysis does not apply to the results of the brainstorming or other plenary sessions. In an attempt to generate a complete list of issues for consideration in the elicitation process, the brainstorming or other plenary sessions were designed to allow maximal interaction among the participants and their results are consequently not independent between panelists. The assumption of *statistical* independence applies only to each panelist's individual responses (mid-values, upper and lower bounds) to the elicitation questions and is used to derive each panelist's uncertainty distribution from their responses. Since the panelists' individual quantitative responses were never discussed in the plenary sessions, the statistical independence assumption is warranted.

It should also be noted that statistical independence of two distributions means that they are uncorrelated, i.e., knowing the value of one does not imply anything about the value of the other. For example, knowing the height of a randomly chosen girl in a school tells us nothing about the height of a second randomly chosen girl. Two distributions can have the same mean or even be identical and still be statistically independent. Thus, even if the mid-values for a particular question were correlated among panelists as a result of panel discussions, there is no reason to believe that the uncertainties about their responses from the individual elicitation sessions were also correlated.

Comment Number: N5-8
Comment: The elicitation of the panelists' individual results and the mathematical manipulations utilized in the aggregation of the individual results is rather complex. There doesn't appear to be a concise discussion relating how the individual panelist's estimates influenced the final results. This makes it difficult to assess the value of the individual panelist's contribution and the aggregate results with regard to LOCA frequency.

Response: Section 7.5 provides individual panelist LOCA frequency estimates for selected LOCA categories. These results illustrate the very large differences between the individual estimates. The use of the geometric mean for aggregation tends to mitigate the contribution of any one panelist to the group estimates of the bottom-line parameters. For the results in this study, this aggregation technique is consistent with the basic objective of obtaining consensus-type group estimates, i.e., estimates which approximate the median of the individual estimates. (See the response to Comment N3-8.) However, the generally large diversity of the individual estimates in this study means that they significantly influence the confidence bounds associated with a given bottom-line parameter estimate.

Comment Number: N5-9
Comment: In the discussion in Section 5.4, "Group Estimates and Confidence Intervals," the statement is made that, "Accordingly, it is assumed that the $\{u_k\}$ values are observations from a lognormal distribution, U." Where $\{u_k\}$, k = 1, ... , n is the set of individual estimates of any one bottom-line parameter of the n panelists. This assumption implies that the panelists' responses are independent and identically distributed. Given the elicitation methodology, which includes anchoring and training, it is difficult to accept the assumption of independence. Similarly, a diverse group of panelists, applying in general different methodologies, would appear to preclude "identically distributed."

Response: It is true that the assumption in Section 5.4 that the individual estimates,$\{u_k\}$, are observations from a lognormal distribution, U, implies that the u_k are independent and identically distributed. This assumption is equivalent to the assumption that the $\{u_k\}$ are samples from a lognormal distribution, U. This is the standard statistical interpretation of a random sample and implies that knowledge of one panelist's distribution provides no information about another panelist's distribution. See the response to Comment N5-7 for additional discussion. Section 5.4 in the final NUREG was clarified to reflect these points.

Comment Number: N5-10
Comment: An assumption that is not well justified or discussed is the fitting into a "split lognormal" distribution the panelists estimates of a lower bound, median and upper bound that is then fit to a lognormal distribution. If a panelist provides his best estimate of these three parameters, he is providing some information regarding his opinion of the shape of the probability distribution. If the panelist realizes that some form of lognormal distribution will be used, this may bias the panelist into picking the lower and upper bound estimates that fit the assumed lognormal shape. If the panelist is not influenced by this, then choosing to fit the lower bound (LB), mid-value (MV), and upper bound (UB) into a lognormal ignores the panelist's judgment of the shape of the distribution.

Response: Contrary to the comment, in providing a mid-value, lower bound and upper bound for an elicited quantity, a panelist did not provide any additional information about the shape of his subjective distribution. It was difficult enough for a panelist to provide the three required percentiles for each question, without having to speculate about the mathematical relationships between the other percentiles of his subjective distribution. Because the panelists had no basis for choosing the shapes of their subjective distributions, they were not asked to speculate on the shapes or to fit their responses to any specific shape. Also, because the panelists were not told that their responses would be fit to split-lognormal distributions, they could not have been biased by this choice as the commenter speculates. See the response to Comments N5-1 and N5-2 for a justification for the lognormal assumption.

Comment Number: N5-11
Comment: The expert elicitation used a process to combine information from the individual panelists. However, it appears that panelists did not provide a consistent set of inputs. For example, on page 4-2, the NUREG indicates that only two panelists predicted base case frequencies beyond 25 years. Another panelist indicated that his base case frequencies are assumed to be largely constant with time, which

appears at odds with the stated scope of the NUREG, which focused on "… primary system side failures that can be exacerbated by material degradation with age." On page xvi, the report notes that not all panelists provided responses for both BWR and PWR reactor types. There is no discussion about the possible impact of the missing information on the final estimates.

Response: It is true that only two of the four base case team members (not panelists) provided base case frequencies beyond 25 years. However, the elicitation panelists, after providing their 25-year estimates, were asked how these estimates would change in the future. The panelists were free to use either their own judgments or use the base case results for anchoring for the future estimates.

Contrary to the comment, there was no preconceived assumption that LOCA frequencies will definitively increase with age. While material aging continues, and by itself, will increase LOCA frequencies, there are mitigation/inspection programs to combat aging. It is therefore possible that frequencies could increase, decrease, or remain constant in the future depending on the effectiveness of these programs. Each panelist provided their assessment about future trends and several expected relatively little increase in frequencies. However, their uncertainties also increased, which causes increases in the mean group frequency. See Section 7.4 for a discussion on the effects of time on LOCA frequency and Appendix L for individual results.

Eight of the 11 panelists who provided responses provided enough information to estimate LOCA frequencies for BWRs and nine of the 11 did so for PWRs. That means that three panelists did not provide information regarding BWRs and two did not provide information regarding PWRs. These were individuals who only had experience with one reactor type, but not both. This self-selection is consistent with the direction given to panelists not to answer questions in technical areas where they had little or no expertise. Information is therefore not missing, but is simply not provided. The only effect is that the sample size of the estimates is reduced which does increase the confidence bounds.

There were cases where a panelist did not initially supply all the information required to calculate individual LOCA frequencies. In these cases, the facilitation team worked with the panelists to provide this information, or to minimize its impact while still allowing final LOCA frequency estimates to be calculated for the panelist. As discussed in Section 3.10 (Final Elicitation Responses), additional assumptions were sometimes necessary to fill in missing information in the panelists' responses. The facilitation team made assumptions or added missing information, whenever possible, in a manner consistent with the philosophy, approach, assumptions, qualitative insights, and other quantitative results already provided by the panelist. The most common reason for augmenting the results was to provide complete coverage intervals for every elicitation response. While panelists often did not provide every coverage interval, they typically did assess coverage intervals in other elicitation responses. If necessary, the facilitation team created the missing coverage intervals to be consistent with these other responses. Any imputed data was subject to panelist review and acceptance prior to final use.

Comment Number: N5-12
Comment: Paragraph 4 of Section 5.1 provides a confusing overview of the uncertainty analysis. The first sentence states that, "It is important that the bottom-line LOCA frequency estimate reflect both individual uncertainty and panel diversity" and is certainly correct. The objective of the methodology is to make a prediction. However, the last sentence in the paragraph states that, "confidence intervals for the estimated bottom-line parameters are developed to reflect panel diversity." Apparently this refers to calculating the mean of the 95th percentile estimates instead of, as in NUREG-1150, combining the full distribution from each panelist. It is just as important to include the individual uncertainty in computation of the confidence interval. As with the selection of the geometric mean instead on the arithmetic mean, this method yields smaller uncertainty bounds and smaller mean frequencies than the NUREG-1150

methods but no justification is provided demonstrating that this method is to be preferred over the NUREG-1150 process.

Response: As stated in Section 5.1, the goal of the study is to estimate the four bottom-line parameters (mean, median, 5^{th} and 95^{th} percentiles) of the LOCA frequencies. The individual estimates of the bottom-line parameters are aggregated to form the group estimates. In the NUREG, "uncertainty" is characterized by the differences between the bottom-line parameters (i.e., the mean and the 5^{th}, 50^{th} and 95^{th} percentiles). If each panelist had no uncertainty about his responses to the elicitation questions, all his bottom-line estimates would be identical and the four group estimates for each LOCA category would also be identical; there would be no uncertainty in each individual estimate. However, the individual estimates might differ, which is what is meant by "panel diversity". The confidence intervals are a measure of panel diversity.

While both the differences in the bottom-line parameters and panel diversity contribute to the uncertainty (in the plain English sense) in the study results, it is important to keep their effects separate. For example, if the mean is adopted as the basis for a regulatory position, then the panel opinion is expressed by the group estimates of the means and the uncertainty in the group estimate is characterized by its corresponding confidence interval. If a conservative estimate of the mean is required, the upper confidence bound can be used.

The geometric mean, rather than the arithmetic mean used in NUREG-1150, was used as the primary method to aggregate the individual estimates to obtain the group estimates because a principal objective of this elicitation study was to determine consensus-type group estimates (see Section 2). In this study, the geometric mean aggregation led to group estimates that approximate the medians of the individual estimates while the arithmetic mean aggregation did not. However, because alternative aggregation methods can lead to significantly different results than those obtain using the geometric mean, a particular set of LOCA frequency estimates is not generically recommended in the NUREG for all risk-informed applications. The purposes and context of the application must be considered when determining the appropriateness of any set of elicitation results.

Sections 5.6.4 and 7.6.4 provide more information on different aggregation schemes and their effect on the bottom-line parameters. See also the response to Comment N5-4 for a related discussion.

Comment Number: N5-13
Comment: There did not appear to be a basis for the method used to adjust panelists' confidence bounds to account for overconfidence. It appears that, in the baseline results, overconfidence adjustments were made to some panelists results but not to others. On page xvii, the NUREG states:

> Ad hoc overconfidence adjustments can result in large, unsupportable increases in the frequency estimates and are discouraged given the extreme skewness of many of the individual distributions. Error factor adjustment was used to increase the low panelists' uncertainties up to the geometric mean of the group. This results in only modest increases to the baseline mean and the 95th percentile.... This adjustment should be performed on the baseline results to account for expected overconfidence.

This indicates that the authors believe that the break frequencies appear overly high when they adjusted all the panelists' input for overconfidence. However, they did adjust the uncertainty estimates for the "low panelists" by calibrating them with overall group average uncertainty. The adjustment of the low panelists estimates is, in turn, deemed acceptable because it results in only "modest" increases in the finale estimates. This appears arbitrary and, to some extent, biased toward maintaining a relatively low frequency estimate.

Response: The error-factor scheme described in Sections 5.6.2.2 and 7.6.2.2 was used to correct those panelists that were more confident about their responses (i.e., they had less uncertainty) than the remainder of the group. This tendency to be overconfident and the appropriateness of overconfidence adjustment in elicitations is discussed in Section 5. Contrary to the comment, the adjustment is applied to panelists with low *uncertainties* (as measured by their error factors), not with low frequencies. Also contrary to the comment, the error factor adjustment did not decrease the frequency estimates, but rather *increased* them (see Table 7.9).

Other types of common overconfidence correction schemes were investigated as part of a sensitivity analysis (See Section 7.6.2). The sensitivity analysis justified the use of the error-factor scheme instead of coverage-interval-adjustment schemes on several grounds. First, the error-factor adjustment varies as a function of the difference between the individual and group (geometric mean) estimates for each parameter. Second, the error-factor adjustment requires no arbitrary determination of which panelists to adjust and the level of the adjustment, in contrast to the coverage interval adjustment. Third, because group uncertainty increases with LOCA size, the error factor adjustment is tailored to increase with LOCA size. This is warranted because larger LOCA sizes deviate further from the operating experience failure and precursor information, thereby making it likely that the degree of overconfidence in estimating frequencies associated with larger LOCAs would increase.

A summary of this justification was incorporated into the Executive Summary of the revised NUREG.

Comment Number: N5-14
Comment: The conclusion drawn in the second paragraph of Section 5.1 does not appear to make sense. The setting described in the paragraph is a matrix whose entries are the estimates of frequencies. The row index identifies, for example, a particular panelist and the column index the particular issue. The objective is to aggregate "the individual panelist responses to obtain group estimates of the bottom-line parameters." The paragraph states that aggregating first over the response index, and then over the panel index, leads to a different bottom-line parameter, than first aggregation over the panel index, and then over the response index. This is a surprising observation. By analogy, changing the order of integration of a multiple integral does not change its value. This appears to illustrate that the results are very dependent on the assumptions and mathematical manipulations performed during the study. The report should explain why this observation is true and why it is acceptable.

Response: The analogy to changing the order of integration in a multiple integral does not apply because the procedures used to calculate the bottom-line parameters from individual panelist responses and aggregate individual estimates to obtain group estimates are highly nonlinear. Accordingly, the order in which these two procedures are conducted leads to different numerical results. The revised NUREG was clarified (Section 5.1) to explain that calculating individual estimates first and then aggregating is consistent with the elicitation structure. Conversely, aggregating the individual responses first and then calculating the bottom-line estimates leads to significant inconsistencies that make it very difficult to either accurately calculate or interpret these estimates.

Comment Number: N5-15
Comment: In the third paragraph of Section 5.2, the statement is made that, "... after choosing appropriate anchoring frequencies, each panelist chose adjustment ratios to relate the anchoring frequencies to the associated LOCA frequencies for that piping system/non-piping subcomponents after 25, 40, and 60 years of plant operation." Assume that the anchoring frequencies are known perfectly, and, therefore, do not contribute to the uncertainty in the prediction. Then the dominant contribution to the prediction of the LOCA frequency over time (25, 40, and 60 years of plant operation) and its uncertainty is the physical phenomenon of degradation of the piping system/non-piping subcomponents. Models of

failure under these conditions take the form of degradation paths over time. To fully capture the degradation phenomenon, the statistical analysis of failure in this context requires the analysis of the degradation paths via a mixed-effects statistical model. PFM, as used by some of the panelists, does not fully capture the underlying error model of the degradation phenomenon, and, thereby, is likely to underpredict the uncertainty.

Response: As indicated in the comment, a significant dominant contribution to the prediction of the LOCA frequency over time (25, 40, and 60 years of plant operation) and its uncertainty is the physical phenomenon of degradation of the piping system/non-piping subcomponents. This was a principal objective that the elicitation procedure was structured to assess. The elicitation specifically required each panelist to provide ratios and uncertainty estimates with respect to chosen base case frequencies, or to choose other anchoring frequencies. While some panelists used PFM models to inform their ratio estimates, they also used their knowledge of the operational experience, various sensitivity analyses, and the differences among the base case results. All of this knowledge combined was used by the panelists to come up with their ratio and uncertainty estimates. No one panelist simply used the result from one PFM study to quantify these measures.

As stated in the NUREG, it is a well-known characteristic of elicitations that panelists tend to underestimate the uncertainty associated with their responses. Elicitation uncertainty results are sometimes adjusted to account for this phenomenon. Sensitivity analyses described in Section 7.6 indicate that some panelists very likely *overestimated* their uncertainties. Therefore, the error factor correction scheme described in Section 7.6.2.2 was developed to adjust only those individual estimates which appeared to underestimate uncertainty while not modifying the other individual estimates.

Comment Number: N5-16
Comment: The prediction of LOCA frequency and its uncertainty in this report tries to capture a priori the effects of in-service inspection. In-service inspection is conditional on the state of degradation. Thus, it is the degradation path over time and its uncertainty that needs to be characterized to fully capture the behavior of the degradation with in-service inspection taken into account, and not only the values at 25, 40, and 60 years. If we now add the uncertainty in the anchoring values, the analysis becomes very muddled. Based on current information, how can one "anchor" predicted values? This appears to be the case with regard to the elicited values at 25, 40, and 60 years.

Response: In-service inspection is just one variable that the panelists needed to consider when providing their estimates. The interrelationship among all of the classes of variables which affect LOCA frequencies (i.e., geometry, loading history, materials, degradation mechanisms, and mitigation & maintenance) is the essence of the elicitation. The anchoring frequencies for the base case scenarios simply provide a starting point for each panelist. Each panelist had to rely on their knowledge of the variable classes above to identify the significant contributing variables and then provide estimates on their effects.

The effects of ISI are implicitly included in the operating experience database because this is often the method used to identify precursor failure through quantifying degradation. The effects of ISI can also be modeled using PFM if inspection intervals and probability of detection (POD) curves as a function of flaw size are available or assumed. Inspection and other mitigation measures are the reason that LOCA frequencies do not increase substantially with service life. Without these measures, simple degradation models would imply that these frequencies increase.

Comment Number: N5-17
Comment: There is a basic premise or assumption stated in the elicitation that the panel responses show no significant systematic bias. There are peer reviewed papers that suggest such an assumption needs to be tested for and confirmed, not merely assumed.

Response: The NUREG does not specifically state that "the panel responses show no significant systematic bias". Rather, Section 3.3 states that "a basic premise in using an expert elicitation process is that the panel responses as a whole have no significant systematic bias" and goes on to list some of the elements of the elicitation procedure which are designed to achieve this goal. The final NUREG was revised to expand this list to include the following:

- Constructing the panel with experts from all relevant technical areas and institutional/organizational affiliations,
- Conducting elicitation training to identify possible sources of bias and conduct an exercise involving "almanac-type" questions with known answers,
- Providing operating experience data and base case scenarios for anchoring and validating responses to the panel,
- Formulating the elicitation questions to avoid response bias, and
- Conducting individual elicitation sessions to eliminate the possibility of group dynamics influencing panelist responses

It should also be noted that the results of the training exercise did not demonstrate any systematic bias by the panel (see Appendix C).

Comments Related to Section 6 of NUREG-1829

None

Comments Related to Section 7 of NUREG-1829

Comment Number: N7-1
Comment: In the NUREG, the charts of break frequency versus threshold break size show six points for the threshold LOCA categories with straight lines connecting the points. Clarify whether the single point values represent equally the frequencies of all breaks within that range of sizes or if the straight lines can be used to determine frequencies of particular intermediate sizes inside the ranges. Would a stair-step function between the points be more appropriate since any range represents a population of piping whose break frequency is not further subdivided by the panelists' input?

Response: This comment is similar to Comments 3-3 and 7-5 from Appendix M of the revised NUREG. See the response to Comment 3-3 from Appendix M for guidance on interpolation schemes.

Comment Number: N7-2
Comment: The expert elicitation process differed in significant ways from the processes used in the well regarded NUREG-1150 and NUREG/CR-5750 elicitations. These differences are underplayed in the elicitation report. Differences include aggregation methods for pulling together panelist results, ...

Response: The assumptions, processes, and approach utilized in NUREG-1829 are thoroughly documented. NUREG-1829 additionally contains additional sensitivity studies (not contained in other documents referenced by commenter) in order to systematically assess differences resulting from different assumptions and analysis techniques. The assumptions, processes, and approach chosen for the NUREG-

1829 estimates (see the Executive Summary) are also justified using physical or statistical rationale, or to ensure that the results are compatible with the framework used in the elicitation. Additionally, a thorough description is provided on how other assumptions or analysis approaches (including arithmetic-mean aggregation) affect the results. These effects are clearly stated in the Executive Summary. See the Executive Summary and Sections 5.6.4 and 7.6.4 for more information.

Comments Related to Section 8 of NUREG-1829

None

Comments Related to Section 9 of NUREG-1829

None

Comments Related to Appendix A of NUREG-1829

None

Comments Related to Appendix B of NUREG-1829

None

Comments Related to Appendix C of NUREG-1829

None

Comments Related to Appendix D of NUREG-1829

Comment Number: ND-1
Comment: In Appendix D, regarding estimates made by B. Lydell using leak precursors, one key assumption made is that larger break frequencies may be extrapolated from smaller break frequencies by a simple power law relationship. However, the validity of this approach is not provided. The smaller break data is very sparse, and there is very little or no data in the larger break regions, which indicates huge uncertainty using any chosen extrapolation method. There may be other more conservative methods of forming extrapolations. The basis for the extrapolations and the assumed uncertainties should be provided.

Response: In response to this comment, Section D.5.2 has been revised to more clearly describe the basis that Bengt Lydell used for calculating conditional weld failure probabilities. In summary, he has evaluated the failure probability as a function of break size for ASME Class 1, 2, 3, and balance of plant piping (i.e., ASME B31.1 piping) to demonstrate that ASME Class 1 (i.e., LOCA sensitive) failures are expected to have a much lower conditional failure (rupture) probability (CFP) than ASME B31.1, Class 2, or Class 3 piping. Additionally, failure information from these other piping classes provides CFP estimates for larger LOCA Categories (than LOCA Category 1). This information is necessary because there have been no large Class 1 piping failures, but there have been large ASME B31.1 piping failures.

A separate two-parameter power law form is then assumed for the CFP to extrapolate the Class 1 condition failure probability relationship from Category 1 LOCAs to higher break sizes for both IGSCC and thermal fatigue mechanisms. The power law shape is consistent with the aforementioned CFP relationships for ASME B31.1 and Class 2/3 piping, as well as the classical CFP based on vibratory fatigue. The power law relationship is calibrated to agree with the operational experience for Class 1 failures in 1 inch and smaller diameter piping. This relationship is then used as a prior constraint on the mean frequency for a Beta distribution. The Beta distribution is then updated based on operational experience as a function of pipe size. This posterior distribution is then applied in the base case analysis to extrapolate the operational-experience leak rate frequency to the other required LOCA sizes.

A more complete description of the approach is provided in revised Section D.5.2.

Comments Related to Appendix E of NUREG-1829

Comment Number: NE-1
Comment: In Appendix E, regarding estimates made by W. Galyean using service history, it is assumed that the LOCA frequency is reduced by ½ order of magnitude for each step increase in size category. However, the validity of this approach is not provided. There are other methods for extrapolation. The basis for the extrapolations and the assumed uncertainties should be provided.

Response: This comment is very similar to Comment N4-1. See the response to this comment for discussion on the appropriateness of this assumption. Additional related information is also provided in the response to Comment N4-8. Sections 3.5.1.2, 4.2, and Appendix E have been modified to reflect the information contained in the response to Comment N4-1.

Comments Related to Appendix F of NUREG-1829

Comment Number: NF-1
Comment: In Appendix F, regarding estimates made by D. Harris using a PFM method, assumptions which use the stresses for the pressurizer surge line elbow from NUREG/CR-6674 are characterized as overly conservative. A "refined" stress distribution based on work sponsored by Electric Power Research Institute (EPRI) (Reference F.6) is also used in the PFM model which greatly reduces the resulting break frequencies. In a previous review of the EPRI analysis, the staff identified numerous issues with the EPRI method and results (reference letter from P. Kuo, Nuclear Regulatory Commission, to F. Emerson, Nuclear Energy Institute, dated January 21, 2004). It is not clear which stress assumption is used in the final estimates. The actual stress distribution in any given elbow and its effect on break frequency is highly uncertain, especially regarding residual and thermal stresses, and it would appear that the most conservative calculated stress distribution should be assumed. The basis for the assumed stress distributions in all of the analyzed components and the basis for the estimated uncertainties should be provided.

Response: Information on the stresses used for the probabilistic analyses is included in Appendix F of the NUREG, with the exception of the refined analysis for the surge line elbow. The stresses used in the surge line evaluation were based on the actual stress analysis for a CE-designed plant developed to address surge-line thermal stratification in response to NRC Bulletin 88-11. The loadings were based on the methods approved by the NRC staff in the CE Owner Group Report CEN 387-NP, "Pressurizer Surge Line Flow Stratification Evaluation," Rev. 1-NP, December 1991. Additional evaluations of the local stress distributions in the elbow were conducted to determine the detailed stress distribution around the circumference of the elbow. The critically stressed location that produced the highest probability of cracking was the circumferential stresses in the side of the elbow due to thermal stratification bending. The detailed stresses are proprietary.

Appendix F of the revised NUREG has been clarified to reflect this response.

Comment Number: NF-2
Comment: In Appendix F, regarding estimates made by D. Harris using a PFM method, an "alternate" procedure is used because the computations are too computer-laborious. The estimates for larger breaks are based on straight line extrapolations on semi-log plots of frequency versus crack length (Figure F.2). However, plots for the PFM computations for smaller cracks have several inflections, which does not indicate the validity of this extrapolation approach. There are no actual data upon which to base such an extrapolation, and there may be other more conservative methods of forming extrapolations. The basis for the 9 extrapolations and the assumed uncertainties should be provided.

Response: The alternative procedure was devised to estimate failure probabilities when direct Monte Carlo simulation was prohibitive. The alternative procedure is described in Section F.3.3.4 of Appendix F and justification for its accuracy is also provided. The alternative procedure generally gives higher (conservative) leak probabilities than direct Monte Carlo simulation in cases where direct comparisons are possible (Tables F.13 and F.28).

Comments Related to Appendix G of NUREG-1829

Comment Number: NG-1
Comment: In Appendix G, regarding estimates made by V. Chapman using a PFM method, the failure probability per year is assumed to equal simply the cumulative probability divided by the number of years of operation. However, for a fatigue crack growth prediction such as that performed, the probability per year is not constant and significantly increases over time. This results in underestimating the probability per year when determined from the cumulative probability divided by the number of years. Also, it is not clear if this was generally done for all of the other panelists' estimates.

Response: A fatigue-based failure mechanism has an increasing hazard function for a single component until it is repaired or replaced. However, the hazard function associated with the entire plant only increases with time if it is assumed that inspection and repair/replacement hazard decreases are less than increases due to global aging. However, taking the integrated failure probability over N years and dividing by N is, by definition, the average failure probability. It is also true that this is not necessarily the failure probability in the last year. Most importantly, the failure probability in the final year cannot be determined from this average failure probability.

In the base case examples, the values used from PRODIGAL runs were the integrated failure probabilities over 25 and 60 years. These were then converted to average failure probabilities simply by dividing by the number of years. For the elicitation results, V. Chapman provided average failure probabilities over the first 25 years, over the next 15 years and over a final 20 years. The first average failure probability comes directly from the PFM model. The average failure probability over the next 15 years is, however, a conditional failure probability. That is to say, this average failure probability is the average failure probability over this 15-year period given the degradation that had accumulated over the first 25 years but failures that occurred in the first 25 years. The same interpretation applies to the calculation of the average failure probability for the final 20-year period. These average failure probabilities correspond to average failure frequencies over the time period of interest.

The elicitation current-day estimates (25 years of operation) were intended to represent current plant conditions and are therefore equivalent to instantaneous LOCA frequency estimates. For example, V. Chapman's responses indicate he believes that the average failure frequencies are similar to the instantaneous values. However, the 40-year and 60-year estimates were not explicitly defined in the elicitation to represent either instantaneous frequencies at 40 and 60 years of operation or averaged frequencies between 25 – 40 years and 40 – 60 years of plant operation, respectively. Because the interpretation of the future time period estimates was not explicitly defined, the panelists were free to assess the relevance of each type of estimate and then provide either averaged or instantaneous frequencies as they deemed most appropriate. Each panelist was also asked to discuss possible differences between averaged and instantaneous frequencies for the future time periods. The lack of an explicit definition, while important for understanding the context of the results, is not of practical significance because future changes in generic LOCA frequencies are generally expected to be gradual. Furthermore, any differences between average and instantaneous frequencies will be dominated by the uncertainty in the estimates.

Comments Related to Appendix H of NUREG-1829

None

Comments Related to Appendix I of NUREG-1829

None

Comments Related to Appendix J of NUREG-1829

None

Comments Related to Appendix K of NUREG-1829

None

Comments Related to Appendix L of NUREG-1829

Comment Number: NL-1
Comment: The NUREG provides box and whisker charts indicating the various individual panelist estimates by a designated letter, but the identity associated with the letter designation is not provided. This information should be provided in order to allow association of the estimates with the methods of obtaining them.

Response: Early in the elicitation process (on the third day of the elicitation kick-off meeting), the subject of confidentiality was raised. The NRC indicated that all information provided as part of this study would remain confidential. Panel anonymity is an essential element in dealing with one of the motivational biases (see Section 3.3.1 of NUREG-1829) that can potentially creep into the elicitation process. Social pressure can lead to 'groupthink' where panelists may suppress their doubts or differing opinions in order to attain consensus. It can also be manifested in a panelist's response to verbal and non-verbal feedback from an interviewer. In addition, a panelist's responses might be influenced by his perception of what might be acceptable to the facilitation team, the other panel members, or the technical community that ultimately reviews the results. Such 'groupthink' may degrade the elicitation process by artificially reducing the uncertainty in the combined group responses. Therefore, it was decided in the beginning of the process that the assumptions, methodology, elicitation results, and calculated LOCA frequencies that stemmed from this study would be reported along with the names, affiliations, and credentials of the expert elicitation panel and the facilitation team. However, no individual reference to individual opinions would be documented in the NUREG.