



RECEIVED

2009 JAN -6 AM 8:38

Rules and Directives
Branch
USNRC

11/13/03

6872-64374

14

E-RIDS = ADM-83
Call = M. Maloy (MSMB)
R. Frabin (RKF)

United States Nuclear Regulatory Commission

Template = ADM-013

These comments are provided for consideration as part of the NRC's annual review of the Reactor Oversight Process.

While I am an NRC staff member I am providing these comments: 1) anonymously, because management statements to the effect that staffers who have criticisms of the ROP just need more risk training sends a message, at least to me, that dissent is unwelcome, 2) publically - because the ROP feedback process, while an adequate mechanism for fixing discrete procedural problems, has not been demonstrated to be an appropriate forum with sufficient management involvement for discussion of more programmatic issues and 3) now - because after four years, the process has been in place a sufficient time to draw conclusions on its performance. Further, while there have been past efforts to address issues with the ROP, they have not taken on/addressed some of the more substantial issues.

These comments represent the views of only one staff member however, many of the issues discussed below are concerns of other staff members as well.

It is understandable that the nuclear industry looks somewhat favorably on this relatively new process because the action matrix makes NRC actions more predictable, as presently implemented there are fewer "significant" actions issued, for those that are issued there are extended opportunities to offer any manner of outside the design basis risk arguments to downgrade the findings, and finally there are very few civil penalties issued to power reactor licensees any more. For the public, the benefits of the new process are less clear. The ROP does provide a predictable response but it has yet to demonstrate measurable improvement over the old SALP system in detecting declining performance. It has also introduced a whole new level of complexity that must be understood to have any confidence in what the NRC is doing. Previously, a member of the public had to understand the technology of the plant and its associated language and acronyms. Now on top of that, a member of the public has to acquire a working knowledge of risk and its associated language/acronyms including when/how issues/events are evaluated using Core Damage Frequency, Large Early Release Frequency, or Conditional Core Damage Probability.

As a staff member I have heard repeatedly the ROP is a "better" process. I would agree that an action matrix approach is an improvement over the SALP and its problem plant list. However, how has the system that produces the inputs to the action matrix been shown to be "better"? Yes, the reactor systems related SDPs are "objective" because they rely on numbers but are they and the other SDPs producing the right results? While the review at Davis Besse is showing us that hard to quantify things like "safety culture" are at the root of many of the issues there, we as an agency are putting more and more trust in a small group of risk trained specialists using tools that don't do well with such unquantifiable issues. Our justification is that problems will manifest themselves as poor performance that we can quantify and we will act on it before a real big problem occurs. Do the action matrix and the SDPs as presently constructed do that? What metric (s) show that to be the case?

Below in no specific order are some areas in which the ROP could be improved:

1. There appears to be a need for some type of validation on the thresholds for Green, White, Yellow and Red findings, both within a given cornerstone and across cornerstones.

a. Within a cornerstone - Very few actions year to year rise to the level of even White significance (low to moderate). That could be because there are few actually significant issues or it could be that some portion of issues found now to be green are in fact more significant. More significant meaning that the risk determination fails/cannot quantify important information or that all the risk is correctly taken into account but the threshold needs to be lower for it to be a valid indicator. The fact that the thresholds for green, white, yellow and red for the numerically based SDPs are the same as they have been since the initial implementation of the program implies that the thresholds were initially and are still appropriate. How do you know they are correct? Said another way, what benchmark(s) are greater than green findings, as now defined and accounted for (multiple whites being a degraded cornerstone etc.), measured against to ensure they are an adequate (early enough) indicator of degraded or declining performance? Can events that have occurred be used to improve our indicators? Might another indicators such as, lowering the white threshold, giving more significance to findings in certain cornerstones, single greater than green findings in multiple cornerstones, or total numbers of findings (including green findings) be equally or more valid predictors of performance? Note: certain SDPs count/have counted issues per time and in effect acknowledge some added significance to otherwise green findings that occur with some frequency.

b. Across cornerstones - A relatively large portion of the greater than green findings issued to date are in areas such as EP and HP. Given that each greater than green finding, regardless of area, affects the action matrix equally it would seem there needs to be proper normalization across areas to ensure that a white in one area is equally reflective of licensee performance as the same color of finding in another area. As each SDP was developed by subject matter experts in that area, was there any effort to normalize similar colored findings across SDPs? More importantly, is such normalization periodically performed or re-performed when, for instance, a given SDP is significantly revised? The enforcement policy specifically cautions "Comparison of significance between activity areas are inappropriate" because of the difficulty there would be in comparing across disciplines. Given that the SDPs focus on outcomes and not causes such as poor oversight or procedural adherence, which could be compared across disciplines, how are findings across areas judged comparable? Are areas such as EP and HP alone valid indicators of overall performance? Are such areas over-represented in the mix of greater than green findings, are plant operational areas under-represented or maybe both? How do you know? Why is it that a plant like Cooper, with a long history of operational performance issues including an extended shutdown to fix hundreds of environmental qualification findings, is only in enhanced oversight because of EP issues?

2. The bands for what constitutes a finding of a particular color appear to overstate the degree of accuracy that is reachable using the tools available. a) The uncertainty/error bands associated with some very thorough but complex determinations may span two or even three colors. In fact, on a given issue, multiple one decade answers may be "right". Are the bands in use too narrow or rigidly defined? b) PRA quality is important and will realize some improvements however it remains to be seen if a standard of quality of PRA "commensurate with its application" is going to bring about quality that is actually good enough to define risk into the narrow bands of the ROP. Further, without standardized PRAs, comparison of results between plants and maybe even between systems

in a given plant is problematic. Results at the E-4 through E-7 levels are being compared and slight differences in HRA assumptions, generic vs plant specific equipment data, and assumed outage times will have significant impacts. Further, given that each plant starts with a unique baseline CDF, how are changes to CDF within given fixed one decade bands the proper measurement? Shouldn't there be some consideration of change relative to the baseline? c) MC 0609A 2.5 recognizes that the external events risk contribution may significantly change the color of the SDP determination, yet the quality and rigor of licensee IPEEEs is more variable than IPEs. How is this additional variation in quality and rigor accounted for in the existing narrow color bands? Doesn't the need for the NRC staff to do "best estimate" calculations when it comes to many issues involving external events acknowledge that many IPEEEs are inadequate? More generally, why PRAs which are now so heavily relied on by the NRC and the licensees are not standardized, docketed, subject to uniform quality standards, and tied to a formal change process has got to be a mystery to many NRC stakeholders.

3. Objective 02.03 of Manual Chapter 0305 states that the ROP assists management in making timely decisions regarding appropriate agency actions to oversee, inspect and assess licensee performance. Is the ROP as now implemented achieving the objective? Specifically, if the outputs of the ROP are in fact suppose to determine agency response to events, how does the agency response to the Indian Point tube failure and the head cracking at Davis Besse, to only mention two events, demonstrate that approach? In those events the NRC properly authorized and expended the resources necessary to address the immediate response to the event. However, the NRC also committed to and took many longer term follow up actions such as follow up team inspections long before either PI or SDP outputs were completed which justified such expenditures. Is assessment driving response or vice versus? Does the commitment of long term resources before ROP determinations are available indicate the process as now constructed and/or implemented is too unwieldy or complex to meet the stated objective?

4. The current MC guidance says that the risk increase of a given performance deficiency is evaluated discretely and not considered in an integrated fashion with other performance deficiencies that affect the same equipment. That approach was criticized externally after it was used for various deficiencies at Davis Besse and the defenses of that approach given by various NRC managers have been in the press. It would seem that an integrated evaluation of all the degraded conditions that affect a system that are known: a) provides a fuller disclosure of the risk and b) avoids the perception that the NRC looks at licensee performance in merely in an event by event fashion when in fact the goal is to have an integrated assessment of performance. That view aside, if the NRC is going to use this deficiency by deficiency approach, it should do it consistently. When the NRC evaluated a degraded fire door condition at ANO, initially it was classified a potential white finding. Subsequently, the licensee argued that a fire watch unrelated and unaware of the deficiency should be credited because that fire watch just happened to transit the area in question during the time period in question. The NRC accepted that argument and because of that, the deficiency received a final green disposition. When are/should surrounding but not directly related circumstances be included in the NRC assessment of a particular issue?

5. The action matrix is an improvement over SALP. However, when a plant reaches the far right of the action matrix or has a significant event the NRC still does not have any guidance for making

a shutdown decision, despite a longstanding recommendation from the GAO in that area. Recognizing that each situation is unique, there can be no exact formula. However, (a) Finding #4 of OIG report 02-03S argues for a better defined process. Despite having concurrence on an order from all the necessary offices and OGC, the Director NRR thought he lacked the regulatory authority to issue the order to Davis Besse. It would seem that an acceptable process would not leave a manager in such a situation but would lay out what the concurrences meant and what methods were available to resolve the question of regulatory authority and (b) without any guidance, the often criticized "show of hands" for Davis Besse would be as valid as any shutdown decision making process in existence. Does the NRC consider the decision-making process used to allow Davis Besse continue to operate with head cracks "a good and appropriate model for future actions" as advocated by Mr. Holahan in his 1/14/03 note to the Commission? If so, how does the NRC answer the critique of the process, in particular the validity of the staff's risk argument, Mr. Long offers in his 2/20/03 note to the Commission? If not in agreement with Mr. Holahan's assessment, what process will the NRC use when confronted with a similar situation in the future?

6. If the ROP and the decision to move away from N+1 resident inspectors have proven effective, why is it that at the end of each of the last couple of years, inspector-qualified personnel from NRC headquarters have been called on to complete the base line (minimum) inspection programs at numerous sites? If the answer is that significant resources were needed to be expended at Davis Besse, why isn't the contingency for such a "problem plant" built into the assumptions, as there always seems to be at least one such plant?

7. Timeliness of the assessment process continues to be an issue despite timeliness goals. It would appear to be because timeliness is not tied to discovery/identification but rather report issuance which may occur long after an issue is first identified. As stated in the ROP implementing documents, a finding is not placed in the action matrix until it is a final determination. Many preliminary SDP assessments are not sent to the licensees for many months after the item occurs/is first identified. Similarly, the final findings are not completed in some cases for many months after the issuance of the preliminary finding, resulting in the actual posting of the final finding lagging the actual occurrence/identification of the event or issue by many months. Is the correct balance of accuracy of the determination and timeliness of issuance being achieved if the action matrix is trying to evaluate current licensee performance?

8. Manual Chapter 0305 Section 0305-1 "Purpose" states that the ROP "integrates the NRC's inspection, assessment and enforcement programs." Given that actions that receive only severity levels under the Enforcement Policy do not appear in/affect the action matrix, how does the ROP in fact fully integrate the enforcement program? Looking forward, if another plant has information accuracy problems like Davis Besse and such problems are detected and dispositioned using enforcement policy, how will inspection resource allocations, which are made through the ROP, be adjusted for these non ROP concerns? Would the staff wait to increase the inspection resources until hardware issues manifest the significance of the information accuracy problem?

9. 0609A states that the SDP is to estimate the increase in CDF due to conditions which contribute "unintended risk caused by licensee performance". It goes on to say that other events that contribute to overall risk fall within "acceptable plant normal operating risk." In other places this

normal operating risk is referred to as baseline risk. a) Occasionally when risk assessments are performed certain risk contributors are excluded because the assertion is made they are part of the normal operating/baseline risk. What contributors can be excluded in such a manner? Besides being all risk not evaluated through an SDP, what exactly is this normal operating or baseline risk? Is it an average value of risk contributions from maintenance, random failures, and equipment outages? Does it include the presence of a certain risk contribution attributed to longstanding but undiscovered design flaws or QA violations? Is it a limiting value arrived at by considering the risk contribution of having the maximum amount of TS required equipment out at a given time? Does the definition differ from PRA to PRA? If so, how are comparisons of risk increases from approaches that define this term differently valid? Example: If a motor driven AFW pump was inoperable due to a performance deficiency would the risk associated with the other motor driven AFW pump being out for maintenance fall into normal operating risk? What about the unavailability of one steam supply to the steam driven AFW pump? Both? What about emergency power to a power operated dump valve? What about all three? b) Given the above guidance on what risk is considered in an SDP, if a licensee performed a 50.65 (a)(4) assessment, determined that the group of normal maintenance activities not due to be performed for some time would increase CDF by E-5, and despite those factors took the equipment out of service to perform the maintenance, this would fall into normal operating risk because it was not "unintended risk" and therefore would not be evaluated using an SDP. There also would be no regulatory violation provided all LCOs were met and the licensee not only assessed the risk but "managed" it as well. However, wouldn't the NRC want to capture the decision by the licensee to increase CDF by so much when it was not actually required? But, neither the SDP or the enforcement process would appear to provide a method for that to be accomplished. So, is the standard of "unintended risk caused by licensee performance" the correct one if minimizing all undue/unnecessary risk is a goal of NRC oversight?

10. Even though not specified in the definition of a performance deficiency, as a matter of practice, the licensee performance at issue in a particular finding has to be current licensee performance (and the definition of exactly what that- how far back you can go and why - is undefined). From one perspective, evaluating "current performance" makes sense, as the action matrix is designed to reflect NRC assessment of current performance. However, if past licensee performance results in unintended risk, it is too old to be considered a "performance deficiency", and does not meet the old design issue criteria, how is that risk evaluated/accounted for? Shouldn't be? Example: What if it was identified by an NRC inspector that many years ago a licensee employee used an inadequate procedure to perform an evolution and the evolution resulted in a risk significant situation? Would it be a performance deficiency despite its age? What if the same employee still worked for the licensee and currently would perform similar duties? What if upon review, the management and employees were different but the same inadequate procedure was in place today? Would it then be considered a "performance deficiency" or only if the procedure was used more currently?

11. MC 0612, 05.04.c allows branch chiefs to disposition as green findings certain issues that do not fit into existing SDPs. How often is this provision used? Are there any reviews done to assure consistency of the types of issues that branch chiefs determine can't be put through the SDPs and their bases for determining such issues are green?

12. How the term "performance deficiency" is implemented can have a large effect on the number of greater than green findings. Example: A plant that has three separate and largely independent procedural problems in the maintenance area. Each instance, after evaluation, has a risk contribution of 1 E-6. If the performance deficiency is defined as a failure to adequately implement maintenance procedures, you end up with a single white issue with an associated risk increase of approximately 3 E-6. However, if in each case the performance deficiency is narrowly defined (failure to adequately implement electrical, mechanical and instrumentation maintenance) the result is three separate white findings and a degraded cornerstone. The ROP guidance is silent on which is the correct approach and is similarly silent on how to deal with related issue of additional examples the licensee may discover when correcting the identified issue(s). Are such additional examples assessed separately? Is the risk contribution of those additional examples combined with that of the initial final finding? If so, what if the additional risk would cause a color change? Can the definition of the performance deficiency simply be expanded to include the examples and no additional finding assessed? If so, does the licensee get an opportunity to review/comment? Alternatively, should findings that the licensee discovers through its corrective action efforts be assessed and considered in the action matrix? Doesn't assessing and considering such issues create a disincentive to finding additional issues?