AUDIT REPORT

Audit of NRC's Oversight of Agreement States' Licensing Actions

OIG-06-A-12 April 14, 2006

All publicly available OIG reports (including this report) are accessible through NRC's Web site at:
http:/www.nrc.gov/reading-rm/doc-collections/insp-gen/

April 14, 2006

MEMORANDUM TO: Luis A. Reyes
Executive Director for Operations

FROM: Stephen D. Dingbaum**/RA/**
Assistant Inspector General for Audits

SUBJECT: AUDIT OF NRC'S OVERSIGHT OF AGREEMENT STATES' LICENSING ACTIONS (OIG-06-A-12)

This report presents the results of the subject audit. Agency comments provided at the exit conference on December 6, 2005, and the agency's written response, dated February 13, 2006, have been incorporated, as appropriate, into this report. Appendix C contains the agency's written comments and Appendix D contains OIG's response.

Please provide information on actions taken or planned on each of the recommendations within 30 days of the date of this memorandum. Actions taken or planned are subject to OIG follow-up as stated in Management Directive 6.1.

We appreciate the courtesies and cooperation extended to us by members of your staff during the audit. If you have any questions or comments about our report, please feel free to contact me on 301-415-5915 or Anthony Lipuma on 301-415-5910.

Electronic Distribution

John T. Larkins, Executive Director, Advisory Committee on Reactor Safeguards/Advisory Committee on Nuclear Waste
G. Paul Bollwerk, III, Chief Administrative Judge, Atomic Safety and Licensing Board Panel
Karen D. Cyr, General Counsel
John F. Cordes, Jr., Director, Office of Commission Appellate Adjudication
Jesse L. Funches, Chief Financial Officer
Janice Dunn Lee, Director, Office of International Programs
Rebecca L. Schmidt, Director, Office of Congressional Affairs
Eliot B. Brenner, Director, Office of Public Affairs
Annette Vietti-Cook, Secretary of the Commission
Luis A. Reyes, Executive Director for Operations
William F. Kane, Deputy Executive Director for Reactor and Preparedness Programs, OEDO
Martin J. Virgilio, Deputy Executive Director for Materials, Research, State and Compliance Programs, OEDO
Jacqueline E. Silber, Deputy Executive Director for Information Services and Administration, and Chief Information Officer, OEDO
William M. Dean, Assistant for Operations, OEDO
Timothy F. Hagan, Director, Office of Administration
Michael R. Johnson, Director, Office of Enforcement
Guy P. Caputo, Director, Office of Investigations
Edward T. Baker, Director, Office of Information Services
James F. McDermott, Director, Office of Human Resources
Corenthis B. Kelley, Director, Office of Small Business and Civil Rights
Jack R. Strosnider, Director, Office of Nuclear Material Safety and Safeguards
James E. Dyer, Director, Office of Nuclear Reactor Regulation
Brian W. Sheron, Director, Office of Nuclear Regulatory Research
Janet R. Schlueter, Director, Office of State and Tribal Programs
Roy P. Zimmerman, Director, Office of Nuclear Security and Incident Response
Samuel J. Collins, Regional Administrator, Region I
William D. Travers, Regional Administrator, Region II
James L. Caldwell, Regional Administrator, Region III
Bruce S. Mallett, Regional Administrator, Region IV

EXECUTIVE SUMMARY

BACKGROUND

In accordance with Section 274 of the Atomic Energy Act of 1954, as amended (AEA), the Nuclear Regulatory Commission (NRC) relinquishes its authority to regulate certain byproduct material to 34 States. These Agreement States are responsible for administering approximately 17,300 materials licenses.

The AEA also mandates that NRC periodically review agreements and actions taken by Agreement States to ensure compliance with the provisions of Section 274 of the Act. NRC established the Integrated Materials Performance Evaluation Program (IMPEP) as the mechanism for overseeing Agreement State programs.

To examine a State's program, the IMPEP review teams evaluate and assign ratings to common and non-common performance indicators. Among the many performance indicators considered, only one specifically relates to evaluating the technical quality of an Agreement States' licensing actions. That is, Common Performance Indicator 4, *Technical Quality of Licensing Actions.*

NRC's IMPEP guidance instructs the review team to evaluate the technical quality of licensing programs based on an in-depth, onsite review of a representative cross-section sample of 10 to 25 licensing actions. Based on the review of the sample, the IMPEP team assigns a rating for Indicator 4.

PURPOSE

The purpose of this audit was to evaluate NRC's oversight of Agreement State licensing actions.

RESULTS IN BRIEF

NRC uses a judgmental sampling method to rate the overall technical quality of an Agreement State's licensing actions. However, while NRC should only apply the conclusions drawn from the sample to those license actions selected, NRC projects the results to the overall licensing program. To project sample results to the entire program and to measure the confidence in those results requires statistical sampling. Because NRC uses judgmental sampling, NRC cannot measure the level of confidence in conclusions about the adequacy of an Agreement State licensing actions to protect public health and safety. Furthermore, without

confidence in ratings about a State's licensing program, NRC cannot attest to the confidence level in overall program ratings on the adequacy of an Agreement State program.

RECOMMENDATIONS

This report makes recommendations to the Executive Director for Operations to: (1) Seek and apply guidance from NRC's statistician, and (2) document the rationale and basis for the sampling methodology used.

OIG ANALYSIS OF AGENCY COMMENTS

OIG shared a draft of the report with NRC management on January 31, 2006. The draft report included a recommendation to apply statistical random sampling techniques in conducting reviews of the Common Performance Indicator 4, *Technical Quality of Licensing Actions*. The February 13, 2006, response from the Deputy Executive Director for Materials, Research, State and Compliance Programs stated that, ". . . although we understand the OIG's conclusions on the NRC oversight of Agreement States' licensing actions as described in the OIG report, the staff continues to disagree with the use of random sampling for this application."

OIG maintains that applying sound statistical methods in selecting licensing actions for evaluating Common Performance Indicator 4 will add value to the existing program. However, OIG recognizes that other sampling approaches are available. Therefore, OIG revised the initial report recommendations to allow for flexibility in establishing and documenting the basis for selecting licensing actions to review.

ABBREVIATIONS AND ACRONYMS

AEA	Atomic Energy Act of 1954, as amended
CFR	Code of Federal Regulations
IMPEP	Integrated Materials Performance Evaluation Program
MD	Management Directive
NRC	Nuclear Regulatory Commission
NSTS	National Source Tracking System
OIG	Office of the Inspector General
RSAO	Regional State Agreements Officer
STP	Office of State and Tribal Programs

[Page intentionally left blank.]

TABLE OF CONTENTS

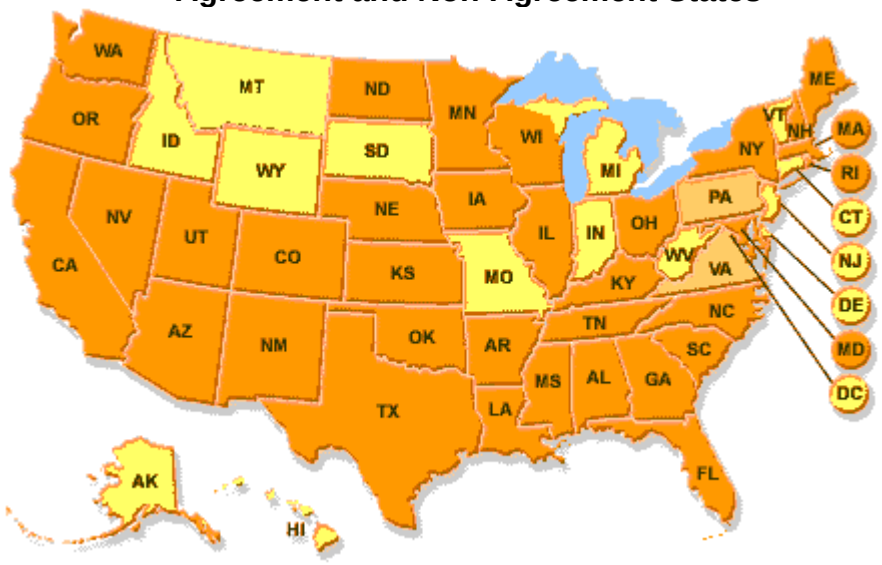
[Page intentionally left blank.]

I. BACKGROUND

The Atomic Energy Act of 1954, as amended (AEA), allows NRC to delegate its authority to regulate certain radioactive material to a State. The mechanism for transferring NRC's authority to a State is an agreement signed by the Governor of the State and NRC's Chairman. The AEA also mandates that NRC periodically review agreements and actions taken by Agreement States to ensure compliance with the provisions of Section 274 of the Act.

NRC's Office of State and Tribal Programs (STP) is responsible for overseeing Agreement State programs. This responsibility includes coordinating the agency's periodic reviews of Agreement State programs to determine their adequacy and compatibility (see Attachment A for NRC's full *Policy Statement on Adequacy and Compatibility*). NRC Figure 1 depicts the distribution of Agreement States and non-Agreement States.

Figure 1
Agreement and Non-Agreement States



Note: Agreement States are shown in orange; non-Agreement States are shown in yellow. Pennsylvania and Virginia are currently non-Agreement States in the process of having their regulatory programs certified for Agreement State status. Source: NRC website [http://www.hsrd.ornl.gov/nrc/home.html] as of April 2006.

NRC's policy is to evaluate its own regional materials programs and Agreement State programs in an integrated manner, using common and non-common performance indicators. Evaluations are intended to ensure that the public health and safety is being

adequately protected and that Agreement State programs are compatible with NRC's programs. The agency conducts these reviews in accordance with established policies and procedures for the Integrated Materials Performance Evaluation Program (IMPEP).

To examine a State's program, the IMPEP review teams evaluate and assign ratings to the performance indicators. The team then makes recommendations to an NRC Management Review Board (the Board) on the overall adequacy of the State's program. The Board makes an overall programmatic assessment on the adequacy and compatibility of each Agreement State program. The assessment is based on, among other things, the IMPEP review team's recommendations, including ratings for common and non-common performance indicators.

Only one performance indicator, among the many considered during an IMPEP review, specifically relates to evaluating the technical quality of Agreement States' licensing actions. That indicator is, Common Performance Indicator 4, *Technical Quality of Licensing Actions*. According to IMPEP guidance for Indicator 4, IMPEP teams should evaluate the technical quality of licensing programs based on an in-depth, onsite review of a representative cross-section of licensing actions. The IMPEP team then assigns a rating for the indicator of—

- satisfactory,
- satisfactory but needs improvement, or
- unsatisfactory.

II. PURPOSE

As part of the Office of the Inspector General's (OIG) *Audit of the Development of the National Source Tracking System* (NSTS) (OIG-06-A-10), OIG examined licensing aspects of NRC's Integrated Materials Performance Evaluation Program. IMPEP is the agency's mechanism for overseeing Agreement State programs.

The overall purpose of the *Audit of the Development of the NSTS* was to determine whether NRC can provide reasonable assurance that licensees use byproduct and sealed source materials safely and account for and control the materials. In accordance with Section 274 of the Atomic Energy Act, as amended, NRC relinquishes its authority to regulate certain byproduct material to 34 States. Agreement States are responsible for administering approximately 17,300 materials licenses. Because Agreement States are an integral part of oversight of byproduct and sealed

source materials, this report focuses on NRC's oversight of Agreement State licensing actions.

See Appendix A for more details on the scope and methodology of this audit.

III. FINDING

NRC uses a judgmental sampling method to rate the overall technical quality of an Agreement State's licensing actions. However, while NRC should only apply the conclusions drawn from the sample to those license actions selected, NRC projects the results to the overall licensing program. To project sample results to the entire program and to measure the confidence in those results requires statistical sampling. Because NRC uses judgmental sampling, NRC cannot measure the level of confidence in conclusions about the adequacy of an Agreement State's licensing actions to protect public health and safety. Furthermore, without confidence in ratings about a State's licensing program, NRC cannot attest to the confidence level in overall program ratings on the adequacy of an Agreement State program.

Sampling Methodology

To have confidence in the accuracy of ratings assigned to Indicator 4, and to assure conclusions projected to the universe (the overall program) based on the sample analysis requires random sampling[1]. Among the factors identified in *Statistics for Management*[2] (Mandel and Laessig) that must be considered in order to determine the appropriate sample size to select when projecting results are—

- the size of the universe,
- the desired confidence in the results, and
- the method of selecting the sample.

Mandel and Laessig state that the sampling methodology and sample size must conform with management's desire for reasonable reliability of the results and confidence in using them.

Conclusions About Licensing Actions are Based on Judgmental Sample

NRC reviews of Agreement State licensing actions yield conclusions for which NRC has no measure of confidence. This is because the method used to select a "representative cross-section"[3] of licensing actions is not statistically sound. Specifically,

[1] Random sampling is a method of selecting a part of the universe by reliance on the law of probability. The purpose of random sampling is to provide representative data for drawing conclusions about the characteristics of the entire universe from which they are selected (Mandel and Laessig, p. 174).
[2] Dr. B.J. Mandel and Dr. Robert E. Laessig, *Statistics for Management*, Dangary Publishing Company, Baltimore, 1996.
[3] Management Directive and Handbook 5.6, *Integrated Materials Performance Evaluation Program (IMPEP)*

staff selects and reviews a judgmental sample[4] of licensing actions upon which they project conclusions about the State's overall licensing actions. IMPEP procedures state that—

- depending on the size of the Agreement State program, the principal reviewer should select between 10 to 25 licensing actions for review;
- these actions should represent a cross-section of the Agreement State's workload, including as many different reviewers and license categories as practical; and
- whenever possible, the selected licenses should include—

  - at least two new licenses,
  - at least three major program amendments (including one denial),
  - at least three license renewals, and
  - at least one license termination or bankruptcy.

IMPEP reviewers use the results from the judgmental sample of licensing actions to draw a general conclusion about the overall technical quality of a State's licensing actions. By projecting the results of the judgmental sample to the State's overall licensing actions, NRC draws a general conclusion when the data refer only to a particular situation.[5]

Specifically, the confidence in projected results based on analysis of a judgmental sample cannot be measured. The main disadvantage of a sample which does not rely on random selection is that when projecting results to the universe there is no way of measuring reliability because ". . . there are no principles which provide a sound basis for computing the sampling error of non-random samples."[6] Additionally, the results of judgmental samples are difficult to defend on grounds of objectivity. Therefore, results based on the judgmental sample of 10 to 25 licensing actions pertain only to the particular licensing actions selected and the results cannot be relied upon to represent the overall performance of the Agreement State's licensing actions.

[4] Samples selected not in compliance with the principle of chance usually are called judgment samples. In this type of sample, the sampler's best judgment, based on past experience, is used in selecting those items for the sample which are believed to give a representative picture of the universe (Mandel and Laessig, pp. 202-203).
[5] Mandel and Laessig, p. 10.
[6] Mandel and Laessig, pp. 202 – 203.

Use of "Smart" Samples

STP staff asserts that using a random statistical sampling methodology could result in issues being overlooked. According to STP staff, IMPEP reviews are performance-based and use a "smart sample" and a team approach that takes advantage of everyone's expertise. The IMPEP teams look at cases that are high risk instead of looking at minor cases such as a licensee requesting the addition of a minor isotope. According to STP staff, "Smart sampling permits the IMPEP team to maximize its focus on licensing actions that approve the use of significant quantities, types and uses of licensed materials that, if not properly licensed and used, could place users and members of the public at greater risk." Additionally, the reviewers' judgment about which licensing actions to select is influenced by interactions between the State and the Regional State Agreements Officers (RSAO) on technical and policy licensing issues. Reviewers' judgment is also influenced by periodic meetings held between the RSAO and Agreement State program management.

The IMPEP program does provide some level of accountability for Agreement State programs by periodically conducting on-site spot checks. However, the confidence in NRC's conclusions about the performance in carrying out licensing actions cannot be measured. This is because they do not include random sampling techniques, and the prescribed sampling methodology lacks a sound basis.

Sample Sizes

IMPEP reviewers select licensing cases on the basis of older procedures already in place and on the institutional experience of its own license reviewers. According to an STP staff member, NRC has not documented the rationale for the sample sizes selected. For instance, NRC has not considered one of the distinct factors identified by Mandel and Laessig: the size of the universe needed to determine a valid sample size.

In fact, the number of licensing actions selected for review is not relative to the State's volume of licensing actions (i.e., the universe). This is because NRC does not know how many licensing actions each Agreement State processes. Specifically, the range of 10 to 25 licensing actions selected for review is the same for all Agreement States. For example, when reviewing licensing actions for the Agreement States California and Rhode Island, reviewers would select 10 to 25 actions for each. However, California is responsible for overseeing 2,077 licenses and Rhode

Island is responsible for overseeing only 58 licenses.[7] The number of licensing actions processed in California presumably would be far greater than the number of licensing actions processed in Rhode Island. According to STP staff, NRC does not have information on the number of licensing actions processed by each Agreement State. An STP staff member said this information is lacking because, under the agreements with NRC, the Agreement States have full authority to run their own program.

Furthermore, NRC limits the sample of licensing actions reviewed to 10 to 25 because of resource constraints. According to an STP staff member, too many resources would be needed to review more than the 10 to 25 licensing actions prescribed by the procedure. However, NRC cannot determine the resources needed until it knows how many licensing actions must be reviewed in each State to produce reliable conclusions about the universe of licensing actions. The STP staff member noted that reviewers do have the latitude to ask management for more resources to complete the review if they see a problem.

Conclusions About Technical Quality of Licensing Actions Lack Confidence

Because IMPEP reviews of the technical quality of licensing actions (Indicator 4) result in conclusions for which NRC cannot measure the confidence level—

- NRC cannot reasonably assure that Agreement State licensing actions are satisfactory to protect public health and safety, and
- Confidence in the Board's conclusions about the overall adequacy and compatibility of an Agreement State program may also be lacking.

According to IMPEP guidance,[8] if a State program is satisfactory but needs improvement for just one or two performance indicators the Board may decide to increase oversight of the Agreement State program. Therefore, while NRC feels confident in their ratings, NRC lacks a basis for measuring confidence in IMPEP ratings for Indicator 4. Without a measure of confidence in ratings for Indicator 4, the confidence in the overall rating of a State's adequacy and compatibility may also be lacking. As a result, the State may not receive the additional oversight warranted.

[7] Number of licenses in this case includes only licenses overseen by the Agreement State program and excludes licenses for which NRC retains regulatory oversight.
[8] Management Directive an Handbook 5.6, *Integrated Materials Performance Evaluation Program (IMPEP)*

CONCLUSION

NRC draws general conclusions about the adequacy of Agreement State licensing actions based on a judgmental sampling methodology. By using a judgmental sample, NRC cannot measure the level of confidence in ratings it assigns to Indicator 4. Without a level of confidence in the ratings assigned to Indicator 4, NRC cannot be assured that the technical quality of Agreement State licensing actions are satisfactory to protect public health and safety.

IV. RECOMMENDATIONS

OIG recommends that the Executive Director for Operations—

1) Seek and apply guidance from NRC's statistician on sampling methodology and criteria for assigning ratings for Common Performance Indicator 4, *Technical Quality of Licensing Actions*; and

2) Document the rationale and basis for the sampling methodology for selecting licensing actions contained in IMPEP guidance for evaluating Common Performance Indicator 4, *Technical Quality of Licensing Actions*.

V. AGENCY COMMENTS

A draft of the report was shared with NRC management on January 31, 2006. The draft report included a recommendation to apply statistical random sampling techniques in conducting reviews of the Common Performance Indicator 4, *Technical Quality of Licensing Actions*. The February 13, 2006, response from the Deputy Executive Director for Materials, Research, State and Compliance Programs stated that, ". . . although we understand the OIG's conclusions on the NRC oversight of Agreement States' licensing actions as described in the OIG report, the staff continues to disagree with the use of random sampling for this application."

OIG maintains that applying sound statistical methods in selecting licensing actions for evaluating Common Performance Indicator 4 will add value to the existing program. However, OIG recognizes that other sampling approaches are available. Therefore, OIG revised the initial report recommendations to allow for flexibility in establishing and documenting the basis for selecting licensing actions to review.

The Agency's response and specific comments are included as Attachment C; and, OIG's analysis and response is included as Attachment D.

[Page intentionally left blank.]

Appendix A

SCOPE AND METHODOLOGY

As part of the OIG's *Audit of the Development of the National Source Tracking System*, OIG examined licensing aspects of NRC's Integrated Materials Performance Evaluation Program, the agency's mechanism for overseeing Agreement State programs. To conduct this audit, we reviewed Section 274 of the Atomic Energy Act, as amended, and the following agency documents:

- Title 10, Code of Federal Regulations (10 CFR), Part 1, *Statement of Organization and General Information*;
- 10 CFR Part 150, *Exemptions and Continued Regulatory Authority in Agreement States and in Offshore Waters Under Section 274*;
- Management Directive (MD) 5.6, *Integrated Materials Performance Evaluation Program (IMPEP)*, and *Handbook*;
- MD 5.9, *Adequacy and Compatibility of Agreement State Programs*;
- STP procedures related to IMPEP, specifically those related to evaluating Common Performance Indicator 4, *Technical Quality of Licensing Actions*;
- *Report of the Working Group on IMPEP Lessons Learned*;
- *IMPEP Good Practices*; and
- Other historical STP documents.

We also reviewed NRC's methodology for evaluating the adequacy of a sample of Agreement State licensing actions and projecting results based on the sample. We compared NRC's sampling methodology against sampling methodology described in the textbook, *Statistics for Management*. Additionally, we discussed the NRC's oversight of the Agreement State Program with an STP manager and staff, and discussed the current IMPEP process and its evolution with an STP staffer.

This audit was conducted in accordance with generally accepted Government audit standards and included a review of management controls related to this audit. This audit was conducted from February 2005 to November 2005.

Major contributors to this report are Anthony Lipuma, Team Leader; Sherri Miotla, Audit Manager; Michael Cash, Technical Advisor; Deb Lipkey, Sr. Management Analyst; and R. K. Wild, Sr. Management Analyst.

[Page intentionally left blank.]

Appendix B

Policy Statement on Adequacy and Compatibility of Agreement State Programs

NRC's policy on compatibility is as follows—

"An Agreement State radiation control program is compatible with the Commission's regulatory program when the State program does not create conflicts, duplications, gaps, or other conditions that jeopardize an orderly pattern in the regulation of agreement material (source, byproduct, and small quantities of special nuclear material as identified by Section 274b. of the Atomic Energy Act, as amended) on a nationwide basis. Compatibility focuses primarily on the potential effects of State action or inaction either on the regulation of agreement material on a nationwide basis or on other jurisdictions. The concept of compatibility does not directly address matters of health and safety within a particular Agreement State; such matters are addressed directly under adequacy. However, many program elements for compatibility may affect public health and safety; therefore, they also may be considered program elements for adequacy. Further, basic radiation protection standards and program elements with transboundary implications, although important for health and safety within the State, should be uniform nationwide for compatibility purposes."

NRC's policy on adequacy is as follows—

"An Agreement State radiation control program is adequate to protect public health and safety if administration of the program provides reasonable assurance of protection of public health and safety in regulating the use of agreement material. The level of protection afforded by the program elements of NRC's materials regulatory program is presumed to be that which is adequate to provide a reasonable assurance of protection of public health and safety. A subset of one of the five elements identified to help provide such reasonable assurance is legally binding requirements addressing protection of public health and safety within the State."

[Page intentionally left blank.]

Appendix C

AGENCY FORMAL COMMENTS

February 13, 2006

MEMORANDUM TO: Stephen D. Dingbaum
Assistant Inspector General for Audits
Office of the Inspector General

FROM: Martin J. Virgilio */RA/*
Deputy Executive director for Materials, Research,
State and Compliance Programs
Office of the Executive Director for Operations

SUBJECT: COMMENTS ON THE FINAL DRAFT AUDIT REPORT OF NRC OVERSIGHT OF AGREEMENT STATES' LICENSING ACTIONS

In response to your memorandum of January 31, 2006, I am providing comments on the Office of the Inspector General (OIG) Final Draft Audit Report of NRC Oversight of Agreement States' Licensing Actions. We appreciate OIG's revising the report to accommodate some of the previous comments the staff provided and discussed with OIG representatives. However, although we understand the OIG's conclusions on the NRC oversight of Agreement States' licensing actions as described in the OIG report, the staff continues to disagree with the use of random sampling for this application.

It is important to note that staff designed the Integrated Materials Performance evaluation Program (IMPEP) such that conclusions on one aspect of the Technical Quality of Licensing Actions performance indicator are not reviewed separately and in isolation from the remaining aspects of this performance indicator, but rather as an integral part of the performance indicator and as integrated into the overall IMPEP assessment. The staff and I continue to support the use of smart (judgmental) sampling for use in the IMPEP process. The use of smart sampling is a highly effective method of assessing small sample lots and has been utilized agency wide for licensing application reviews, integrated safety analysis reviews and inspections, such as for quality assurance, and material control and accounting. The results of smart sampling can be used effectively to qualitatively assess the performance of an organization. This methodology supports NRC's risk-informed and performance-based approach of regulation by reviewing relatively small sample lots of data to make a determination on protection of public health and safety.

Finally, the staff agrees with OIG that scientific sampling has merit when evaluating or auditing certain programs with sufficiently large sample lots. However, the staff and I believe that the use of completely random sampling, with quantified levels of confidence based only on licensing actions, used in the multifaceted performance-based approach of IMPEP, would provide less meaningful information than the holistic approach and, in some cases, could yield results that are misleading. I have enclosed additional specific comments on the draft report for your consideration.

Enclosure:
As stated

SPECIFIC COMMENTS ON THE
FINAL DRAFT OIG AUDIT REPORT OF
NRC OVERSIGHT OF AGREEMENT STATES' LICENSING ACTIONS

1. Page 4, first paragraph. The first sentence is inaccurate and should be changed to read as follows:

   Among the indicators considered during an IMPEP review, Common Performance indicator 4, Technical Quality of Licensing Actions, is focused on evaluating the technical quality of Agreement States' licensing actions in connection with the other performance indicators to determine an overall program performance.

2. Page 4, Results. The audit report concludes that the methodology used by NRC to select a sample of licensing actions to review during IMPEP reviews is not sound. The report goes on to assert that NRC cannot qualify the level of confidence it has in Agreement State licensing actions because of the use of a judgmental sample rather than a scientific random sample. The staff believes that the net result claimed in the Office of the Inspector General (OIG) report is incorrect, i.e., that NRC may not be assured that the technical quality of Agreement State licensing actions are satisfactory to protect public health and safety and that NRC cannot attest to the confidence level in the overall program ratings on adequacy. The use of a scientific random sample would be appropriate for probability sampling of a homogeneous population of items and where a statistical confidence level is required as in American Society for Testing and Materials (ASTM) E 105 - 58, "Standard Recommended Practice for Probability Sampling of Materials." The staff did not design IMPEP to include probability sampling and a calculated statistical confidence level, and to do so would be inappropriate for the evaluation of an Agreement State or NRC Regional program. Probabilistic sampling and statistical confidence levels rely on the use a data range of at least 30 to 40 data points for analysis. IMPEP audits do not typically contain enough data in any one area to yield meaningful statistical interpretations.

   The assumption that the term "representative, cross-section" reference in both Management Directive (MD) 5.6 and the staff procedures refers to a statistical random sample is incorrect. The staff intended to perform what OIG refers to as a judgmental sample to check licensing actions as a significant part of the IMPEP reviewers' and team leader's judgement on the adequacy of the Agreement State licensing program, but not intended to be used as the only evaluation input. The MD intent and procedure requirement is to select a representative sample of the population (licensing actions) in order to support a reasonable or adequate assurance of appropriate regulatory safety in these activities. The use of a scientific random sample in this case is not required, was not intended, and would, in some cases, not be risk-informed or representative of the population being sampled. The staff approach to sampling of licensing actions conforms with the agency practices, and with the requirements of the MD and procedures.

ENCLOSURE

The staff uses a judgmental approach to categorizing activities to be reviewed into subgroups of similar actions. A sample lot of actions is then randomly selected and reviewed by staff to assess the performance of Agreement State personnel in adequately accomplishing these actions within appropriate processes.

3. Page 5, Sampling Methodology. The staff recommends that this section be revised, since it was not the intent of the IMPEP process to conduct a statistical random sample of licensing actions, generate a statistically calculated confidence level, and project the conclusions on overall program performance. A representative sample is intended to be selected, based on the staff's knowledge and judgement of the specific organization's licensing actions, the risk-significance of the licensing actions and previous experience. This approach is consistent with agency practice for licensing application reviews, integrated safety analyses reviews and inspections, such as for quality assurance, and material control and accounting. Most authoritative statistical references, including Juran's Quality Control Handbook, Shewhart's Economic Control of quality of manufactured Product, and B. J. Mandel's Statistics for Management, emphasize that whether a sample is to be used for process control, or a calculated confidence statement, or lot acceptance, or technical or management evaluation or judgement, the sample must be representative and based on knowledge of the population and/or processes that produce them. Many of the industry consensus standards used by the nuclear industry do not require scientific random samples, much less a calculated confidence limit. The American Society of Mechanical Engineers (ASME) Boiler and Pressure Vessel Code sections generally require sampled population. This code is used for many nuclear and other industry safety-related and critical components and systems. ASME Code requirements for In-Service inspection (ISI) specify one of every five welds be inspected. This is basically a process control sample, based on earlier qualification of the process. ASTM standards and specifications in many cases do not require a random sample. Many ASTM standards specifically require a representative sample, e.g., ASTM C-776, Section 5.

4. Pages 5 and 8, Conclusions About Licensing Actions Based on Judgement Samples. The staff recommends that this section be revised to reflect the full context of the OIG/Office of State and Tribal Programs discussions and practices and offers the following text for your consideration:

   The IMPEP team, in its evaluation of the overall Agreement State's performance, seeks to best understand the root cause of why a weakness is identified in a particular performance indicator, whether that indicator is licensing actions, performance of inspections, or any other aspect of the Agreement State's radiation control program. Thus, it may be the case that a weakness discovered in a licensing action is attributable to insufficient training or shortage of personnel in the department or incompatible regulations. The staff believes that these linkages are necessary and would not be revealed by an approach limited to a scientific random sample.

   The NRC staff believes the value added from the integration and interconnection of the five common performance indicators on which the overall Agreement State performance is based is an efficient and effective way to conduct the Agreement State evaluations. The considerable interaction with the various IMPEP team

members (including an Agreement State member) draws on all sources of knowledge available to staff in order to make the best and most accurate assessment of the performance of an Agreement State's program. The emphasis is on a holistic and integrated performance approach to program evaluation to avoid a segmented approach of considering the common performance elements one by one.

As a practical matter, what the team members seek to accomplish under the leadership of an experienced IMPEP team leader, is to bring to bear focus on known or potential weaknesses in the Agreement State program under review. This approach is followed in order to be as maximumly effective as possible in the selection of licensing actions for in-depth review by the team (smart sampling) as one aspect of examination of the questionnaire, review of applicable statues and regulations, analysis of quantitative information of licensing data, review of procedures, and staff and management interviews. Smart sampling permits the IMPEP team to maximize its focus on licensing actions that approve the use of significant quantities, types and uses of licensed materials that, if not properly licensed and used, could place users and members of the public at greater risk. Random sampling of licensing actions will likely result in the identification of a significant number of actions that only authorize the use of licensing materials with low risk since they represent the largest number of licensees.

5. Page 8, Sample Size. This section raises the size of the sample of licensing actions reviewed in the IMPEP process. The OIG report notes that the sample size may range from 10-25 typically and does not account for the number of licensees from large States such as California or Texas to small States such as New Hampshire or Rhode Island. The staff believes that the OIG report failed to properly credit the construction of the sample to emphasis a spectrum of different types of licensees depending on the activities in the State. The staff uses a judgmental approach to categorize similar actions into subgroups. The actions to be reviewed are then selected by staff on a somewhat random basis for further evaluation. Through the use of high knowledgeable personnel, the staff believes that it is able to target possible weakness in the Agreement State's licensing program, which might otherwise be under-represented by using a random sampling approach. The staff's selection is based on a risk-informed and performance-based evaluation in areas of potential weakness.

6. Pages 9 and 10, Confidence in Conclusions About Technical Quality of Licensing Actions Lacking and CONCLUSIONS. As noted above in items 1 through 5, the staff disagrees with OIG's conclusions on the IMPEP process.

7. Page 10, Recommendations. Based upon the comments above, the staff disagrees with these recommendations and suggests that both recommendations be replaced with the following revised recommendation:

   Review and document the rationale and basis for selection of a representative, cross-section of health and safety significant licensing actions in IMPEP reviewer guidance in support of the smart sampling conducted by the knowledgeable IMPEP team.

Appendix D

DETAILED ANALYSIS OF COMMENTS

Where the Agency commented that OIG is incorrect, OIG disagrees and has reaffirmed the accuracy of its statements. The remainder of the Agency's comments provides their opposing stance on the appropriateness and the applicability of the use of statistical sampling in reviewing the technical quality of licensing actions in an Agreement State.

OIG's central message in the report is that, if the Agency desires to have a level of confidence in projecting results of an evaluation based on a sample of licensing actions (i.e., assigning a rating), statistical sampling methods must be used. As mentioned in the report, judgmental sampling provides no way of measuring the reliability of whether the results from the sample give a representative picture of the whole (the whole, in this case, would be the overall technical quality of all licensing actions processed in an Agreement State). While other sampling methodologies exist, statistical sampling provides a reliable and conservative approach appropriate for a regulatory agency with a mission to protect the public health and safety. Therefore, to allow for flexibility in establishing and documenting the basis for selecting licensing actions to review, OIG revised the report recommendations.

A point by point analysis of the Agency's comments follows below.

NRC Specific Comment 1

1. Page 4, first paragraph. The first sentence is inaccurate and should be changed to read as follows:

> Among the indicators considered during an IMPEP review, Common Performance indicator 4, Technical Quality of Licensing Actions, is focused on evaluating the technical quality of Agreement States' licensing actions in connection with the other performance indicators to determine an overall program performance.

OIG Response

OIG disagrees that the sentence is inaccurate. However, to clarify, OIG revised the report to state, "Only one performance indicator, among the many considered during an IMPEP review, specifically relates to evaluating the technical quality of Agreement States' licensing actions. "

Specific Comment 2

2. Page 4, Results. The audit report concludes that the methodology used by NRC to select a sample of licensing actions to review during IMPEP reviews is not sound. The report goes on to assert that NRC cannot qualify the level of confidence it has in Agreement State licensing actions because of the use of a judgmental sample rather than a scientific random sample. The staff believes that the net result claimed in the Office of the Inspector General (OIG) report is incorrect, i.e., that NRC may not be assured that the technical quality of Agreement State licensing actions are satisfactory to protect public health and safety and that NRC cannot attest to the confidence level in the overall program ratings on adequacy. The use of a scientific random sample would be appropriate for probability sampling of a homogeneous population of items and where a statistical confidence level is required as in American Society for Testing and Materials (ASTM) E 105 - 58, "Standard Recommended Practice for Probability Sampling of Materials." The staff did not design IMPEP to include probability sampling and a calculated statistical confidence level, and to do so would be inappropriate for the evaluation of an Agreement State or NRC Regional program. Probabilistic sampling and statistical confidence levels rely on the use a data range of at least 30 to 40 data points for analysis. IMPEP audits do not typically contain enough data in any one area to yield meaningful statistical interpretations.

The assumption that the term "representative, cross-section" reference in both Management Directive (MD) 5.6 and the staff procedures refers to a statistical random sample is incorrect. The staff intended to perform what OIG refers to as a judgmental sample to check licensing actions as a significant part of the IMPEP reviewers' and team leader's judgement on the adequacy of the Agreement State licensing program, but not intended to be used as the only evaluation input. The MD intent and procedure requirement is to select a representative sample of the population (licensing actions) in order to support a reasonable or adequate assurance of appropriate regulatory safety in these activities. The use of a scientific random sample in this case is not required, was not intended, and would, in some cases, not be risk-informed or representative of the population being sampled. The staff approach to sampling of licensing actions conforms with the agency practices, and with the requirements of the MD and procedures.

The staff uses a judgmental approach to categorizing activities to be reviewed into subgroups of similar actions. A sample lot of actions is then randomly selected and reviewed by staff to assess

the performance of Agreement State personnel in adequately accomplishing these actions within appropriate processes.

OIG's Response

OIG stands by the statement that the methodology used by NRC to select a sample of licensing actions to review during IMPEP reviews is not sound. NRC has neither provided OIG with a documented basis nor has it described a sound basis that could provide assurance that conclusions drawn (i.e., ratings) about the State's overall licensing actions (i.e., ratings for Common Performance Indicator 4) accurately reflect the overall performance of the State's licensing actions and that the results are verifiable. Without a documented, sound basis for selecting a "representative, cross-section" of licensing actions, NRC cannot have assurance that ratings for performance indicator 4, "Technical Quality of Licensing Actions," are valid.

With regard to the Agency's statement that IMPEP was not designed to include probability sampling, NRC did design the IMPEP Common Performance Indicator 4 to yield a broad determination of a State's performance in processing licensing actions. However, the attributes of the program (i.e., the use of judgmental sampling) do not allow for such broad determination. In other words, without using random sampling methodology, conclusions drawn about the technical quality of a State's licensing actions (i.e., a "satisfactory" rating) can only apply to the specific licenses reviewed and cannot be projected to the overall licensing actions performed by the State. To make such broad determinations without reviewing 100 percent of licensing actions or using a random sampling methodology is subjective and arbitrary.

OIG disagrees with the Agency's assertions that it would be inappropriate for the evaluation of an Agreement State or NRC Regional Program to include probability sampling and a calculated statistical confidence level and that IMPEP audits do not have enough data points in any one area. The Agency has not consulted with a statistician to assess how or whether statistical sampling methodology could be applied to add value to the process.

The Agency staff also asserts that they intended to perform "judgmental samples" to support a "reasonable or adequate" assurance of regulatory safety and that scientific random sampling is not required. While OIG understands that staff intended to perform judgmental sampling, as mentioned in the report, this technique provides no way of measuring its

reliability. In judgmental samples, ". . . the sampler's best judgment based on past experience is used in selecting those items for the sample which are believed to give a representative picture of the universe. In some instances, where good judgment is used, or by good luck, such a sample may give a fairly accurate representation of the universe. The main disadvantage of a sample which does not rely on random selection is that when the estimate is made there is no way of measuring its reliability, since there are no principles which provide a sound basis for computing the sampling error of non-random samples" (Mandel and Laessig, p. 202).

While OIG acknowledges that scientific random sampling is not specifically required in NRC's IMPEP guidance, assigning a rating (e.g., satisfactory, unsatisfactory) to the Common Performance Indicator 4 during an IMPEP review based on a sample necessitates consideration for assessing the level of assurance in the rating. In other words, given the current methodology, what assurance can NRC give the public that a rating of "satisfactory" is a true representation of the universe?

Finally, with regard to the term "representative, cross-section," no language is included in the report that refers to the term as a statistical random sample. Nonetheless, the term is used in MD 5.6, *Integrated Materials Performance Evaluation Program (IMPEP)*, and *Handbook* and other IMPEP guidance without a specific definition. Consequently, it is open to interpretation. Because judgmental samples cannot be relied upon to give a "representative" picture of the universe (Mandel and Laessig, p. 202), it is easy to see how one might interpret the term to mean a statistical random sample. Additionally, the term "sampling" is often used to describe the selection of a part of a whole field for the purpose of drawing conclusion about the entire universe from a study of a part (the sample). This description is consistent with what the agency's guidance prescribes. The basic requirement of sampling is that a sample be "representative" or serve as an off-spring from the parent or the entire universe from which it is selected. The best way of meeting this requirement is to use a random sample that also is large enough to provide a reasonably accurate picture of the universe (Mandel and Laessig, p. 174).

NRC Specific Comment 3

3. Page 5, Sampling Methodology. The staff recommends that this section be revised, since it was not the intent of the IMPEP process to conduct a statistical random sample of licensing actions, generate a statistically calculated confidence level, and project the conclusions on overall program performance. A representative sample is intended to be selected, based on the staff's knowledge and judgement of the specific organization's licensing actions, the risk-significance of the licensing actions and previous experience. This approach is consistent with agency practice for licensing application reviews, integrated safety analyses reviews and inspections, such as for quality assurance, and material control and accounting. Most authoritative statistical references, including Juran's Quality Control Handbook, Shewhart's Economic Control of quality of manufactured Product, and B. J. Mandel's Statistics for Management, emphasize that whether a sample is to be used for process control, or a calculated confidence statement, or lot acceptance, or technical or management evaluation or judgement, the sample must be representative and based on knowledge of the population and/or processes that produce them. Many of the industry consensus standards used by the nuclear industry do not require scientific random samples, much less a calculated confidence limit. The American Society of Mechanical Engineers (ASME) Boiler and Pressure Vessel Code sections generally require sampled population. This code is used for many nuclear and other industry safety-related and critical components and systems. ASME Code requirements for In-Service inspection (ISI) specify one of every five welds be inspected. This is basically a process control sample, based on earlier qualification of the process. ASTM standards and specifications in many cases do not require a random sample. Many ASTM standards specifically require a representative sample, e.g., ASTM C-776, Section 5.

OIG's Response

OIG acknowledges that the staff did not intend to conduct a statistical sample. However, Common Performance Indicator 4 **is designed** to yield a broad determination of a State's performance in processing licensing actions, yet the attributes of the program (i.e., the use of judgmental sampling) do not allow for such broad determination. In other words, without using random sampling methodology, conclusions drawn about the overall performance on the technical quality of a State's licensing actions (i.e., a "satisfactory" rating) can only apply to the specific licenses reviewed and cannot be projected to the overall licensing actions performed by the State.

As the authoritative statistical references mentioned above state, the sample must be representative and based on knowledge of the population and/or processes that produce them. Similarly, as stated in OIG's response to comment 1 above, the basic requirement of sampling is that a sample be "representative" or serve as an off-spring from the parent or the entire universe from which it is selected. The best way of meeting this requirement is to use a random sample that also is large enough to provide a reasonably accurate picture of the universe (Mandel and Laessig, p. 174). Without some consideration of statistical sampling, the agency has no way of assuring that the judgmental sample provides a reasonably accurate picture of the universe (i.e., is truly "representative").

NRC Specific Comment 4

4. Pages 5 and 8, Conclusions About Licensing Actions Based on Judgement Samples. The staff recommends that this section be revised to reflect the full context of the OIG/Office of State and Tribal Programs discussions and practices and offers the following text for your consideration:

> The IMPEP team, in its evaluation of the overall Agreement State's performance, seeks to best understand the root cause of why a weakness is identified in a particular performance indicator, whether that indicator is licensing actions, performance of inspections, or any other aspect of the Agreement State's radiation control program. Thus, it may be the case that a weakness discovered in a licensing action is attributable to insufficient training or shortage of personnel in the department or incompatible regulations. The staff believes that these linkages are necessary and would not be revealed by an approach limited to a scientific random sample.

The NRC staff believes the value added from the integration and interconnection of the five common performance indicators on which the overall Agreement State performance is based is an efficient and effective way to conduct the Agreement State evaluations. The considerable interaction with the various IMPEP team members (including an Agreement State member) draws on all sources of knowledge available to staff in order to make the best and most accurate assessment of the performance of an Agreement State's program. The emphasis is on a holistic and integrated performance approach to program evaluation to avoid a segmented approach of considering the common performance elements one by one.

As a practical matter, what the team members seek to accomplish under the leadership of an experienced IMPEP team leader, is to bring to bear focus on known or potential weaknesses in the Agreement State program under review. This approach is followed in order to be as maximumly effective as possible in the selection of licensing actions for in-depth review by the team (smart sampling) as one aspect of examination of the questionnaire, review of applicable statues and regulations, analysis of quantitative information of licensing data, review of procedures, and staff and management interviews. Smart sampling permits the IMPEP team to maximize its focus on licensing actions that approve the use of significant quantities, types and uses of licensed materials that, if not properly licensed and used, could place users and members of the public at greater risk. Random sampling of licensing actions will

likely result in the identification of a significant number of actions that only authorize the use of licensing materials with low risk since they represent the largest number of licensees.

OIG's Response

As this report states, "OIG examined licensing aspects of NRC's IMPEP program. Specifically, this report focuses on NRC's oversight of Agreement State licensing actions." OIG acknowledges that Indicator 4 is one of many considered during an IMPEP review. Additionally, OIG understands that if the IMPEP team identifies an issue under Indicator 4, the cause for the issue may be exposed while reviewing another indicator. However, IMPEP teams follow procedures specific to each performance indicator. The message in the report speaks specifically to the procedures followed by IMPEP teams to review Indicator 4. That is, whether or not the technical quality of licensing actions completed by the State are "Satisfactory."

NRC Specific Comment 5

5. Page 8, Sample Size. This section raises the size of the sample of licensing actions reviewed in the IMPEP process. The OIG report notes that the sample size may range from 10-25 typically and does not account for the number of licensees from large States such as California or Texas to small States such as New Hampshire or Rhode Island. The staff believes that the OIG report failed to properly credit the construction of the sample to emphasis a spectrum of different types of licensees depending on the activities in the State. The staff uses a judgmental approach to categorize similar actions into subgroups. The actions to be reviewed are then selected by staff on a somewhat random basis for further evaluation. Through the use of high knowledgeable personnel, the staff believes that it is able to target possible weakness in the Agreement State's licensing program, which might otherwise be under-represented by using a random sampling approach. The staff's selection is based on a risk-informed and performance-based evaluation in areas of potential weakness.

OIG's Response

The description contained in the report of how NRC staffs construct the sample is taken from the State and Tribal Programs' Procedure SA-104, *"Reviewing Common Performance Indicator #4, Technical Quality of Licensing Actions"*. Additionally, OIG believes that by using sound statistical methodology to construct the sample, NRC can achieve a risk-informed sample through stratifying the sample.

NRC Specific Comment 6

6. Pages 9 and 10, Confidence in Conclusions About Technical Quality of Licensing Actions Lacking and CONCLUSIONS. As noted above in items 1 through 5, the staff disagrees with OIG's conclusions on the IMPEP process.

OIG's Response

See response to Specific Comments 1 through 5 above.

NRC Specific Comment 7

7. Page 10, Recommendations. Based upon the comments above, the staff disagrees with these recommendations and suggests that both recommendations be replaced with the following revised recommendation:

Review and document the rationale and basis for selection of a representative, cross-section of health and safety significant licensing actions in IMPEP reviewer guidance in support of the smart sampling conducted by the knowledgeable IMPEP team.

OIG's Response

OIG intends for this report to add value to IMPEP by recommending that NRC consult with a statistician about constructing a sample that will provide a level of assurance in ratings IMPEP reviewers assign to Common Performance Indicator 4. To allow for flexibility in establishing and documenting a sound basis for selecting licensing actions to review, OIG has revised the recommendations to require consultation with the Agency's statistician rather than prescribing the use of statistical random sample.