The National Association for Regulatory Administration

Understanding and Achieving Strong Interrater Reliability

The *Best Practices for Human Care Regulation*, under 2.5 Quality Assurance for Improvement and Control recommends the use of an interrater reliability process.

Interrater reliability means the degree of consistency with which regulatory oversight agencies and individual regulators measure and determine compliance with regulatory requirements.

Strong interrater reliability means that the regulatory oversight agency's staff are measuring and determining compliance using the same methods and coming to the same conclusions about compliance.

Weak interrater reliability means that the agency's regulators are not applying the same methods and/or not coming to the same conclusions.

The 5 Tenets of Interrater Reliability

1) Strong interrater reliability is consistent with NARA's core principles. Strong interrater reliability ensures fairness, objectivity, consistency, reasonableness, and appropriate use of authority in regulatory administration. These concepts are consistent with core principles promoted and encouraged by NARA since the Association's founding over 40 years ago:

- Regulatory administration is to be conducted in a fair and impartial manner, undistorted by personal feelings or biases.
- Regulatory administration is not to be conducted in a manner that is capricious or oppressive.
- Authority in regulatory administration must always be applied with complete fairness and objectivity; authority is always suspect when it is employed without fair, complete, and factual findings.
- The provision of technical assistance to licensees in understanding regulations is an essential component of regulatory administration.

2) Strong interrater reliability minimizes accusations that regulators do not consistently apply and interpret the regulations. The most common complaint about regulatory administration is that regulators are inconsistent in regulatory interpretation and application. A regulatory oversight agency's ability to demonstrate strong interrater reliability is an effective deterrent to such complaints.

3) Strong interrater reliability does not necessarily mean that regulatory interpretation and application is accurate. A regulatory oversight agency where no violations are identified on all or most of the inspections conducted may well have strong interrater reliability, but the absence of violations should raise questions about the effectiveness of the agency's inspection processes. Conversely, a regulatory oversight agency where every inspection finds a high number of violations may also have strong interrater reliability but could indicate unbalanced authority in the regulatory administration process.

4) Strong interrater reliability does not mean 100% is achieved. Regulatory oversight agencies should strive to achieve the highest possible degree of interrater reliability while recognizing that an ideal state of "perfect" interrater reliability cannot be achieved. Regulations that reference "hazards," "cleanliness," etc. require qualitative measurement on a case-by-case basis. It is impossible to list every condition that could be "hazardous" or "unclean," which will inevitably lead to different interpretations by regulators. Additionally, regulators are human who will make mistakes and should not be punished for accidental inconsistencies.

5) Strong interrater reliability processes must include tools, resources, and practices for implementation by regulators. Declarations of advocacy, adherence, and dedication to strong interrater reliability alone are not sufficient. Regulatory oversight agencies must provide regulators with practical resources in order to achieve strong interrater reliability.

Evaluating Interrater Reliability

Regulatory oversight agencies should evaluate the strength of their interrater reliability. Elements of the regulatory administration process used to measure interrater reliability include:

- **The agency's understanding of and approach to interrater reliability.** While most regulatory oversight agencies are familiar with accusations of inconsistent application of rules, many agencies are unfamiliar with the generally accepted practices of interrater reliability. This often leads to the implementation of ineffective corrective measures such as:

  - Conducting "look-behind" inspections (inspections where supervisory staff complete a second inspection of the same licensee shortly after the "official" inspection). Look-behind inspections do not produce reliable results as the setting and circumstances may not be the exact same.

  - Establishing "zero tolerance" policies on inconsistency. In addition to the problems described in Tenet 4 above, a commitment to 100% interrater reliability creates false expectations among licensees; and may significantly decrease regulator's morale.

  - Promoting unnecessary changes to current regulations. While regulations should be revised and updated on a regular basis, regulatory changes are not an effective means to improve interrater reliability; in fact, the adjustment to changes in regulatory requirements usually impairs interrater reliability for months or even years after the changes are made.

  Regulatory oversight agencies should evaluate their understandings of and approaches to interrater reliability to ensure that current practices – or lack thereof – are not inadvertently compromising interrater reliability.

- **The numbers of violations identified during inspections where all regulations are measured by regulator and geographic location.** Licensed settings present on a spectrum of compliance in that a small number of settings are highly compliant, a small number of settings struggle with compliance, and the remainder of settings fall somewhere between these extremes. As a result, it is expected that the distribution of numbers of violations found during inspections where all regulations are measured will be consistent by regulator and geographic location. Inconsistencies in the distribution of numbers of violations identified by regulator and geographic location are powerful indicators of weak interrater reliability.

  Regulatory oversight agencies should measure and analyze numbers of violations identified by regulator and geographic location to ensure a consistent distribution of numbers of violations found during inspections where all regulations are measured.

- **The types of violations identified during inspections where all regulations are measured by regulator and geographic location.** As is the case with numbers of violations identified during inspections where all regulations are measured, it is expected that the types of violations found during inspections where all regulations are measured will be consistent by regulator and geographic location. For example, regulatory violations relating to supervision of persons in case should generally be similar regardless of regulator or geographic location. Inconsistencies in the types of violations identified by regulator and geographic location are powerful indicators of weak interrater reliability.

  Regulatory oversight agencies should measure and analyze types of violations identified by regulator and geographic location to ensure consistency in violations found during inspections where all regulations are measured.

Achieving Strong Interrater Reliability

Regulatory oversight agencies can take steps to maximize interrater reliability. These include, but are not necessarily limited to:

- **Development and use of interpretive guides.** Sound interpretation guidelines and procedure manuals help both licensees and regulators apply regulatory requirements correctly, fairly, and consistently. An interpretive guide should include:

  - The regulation exactly as it appears in the promulgated statutes and regulations;
  - A "Discussion" section that provides clarification about the requirement, addresses frequently occurring situations, and situations that may result in a violation of the regulation;
  - An "Inspection Procedures" section that describes how regulators will measure compliance with the regulation; and
  - A "Primary Benefit" section that includes the benefits of the regulation. "Primary Benefit" is used in lieu of "Intent of Regulation" as regulations may have more than one intent.

  Regulatory oversight agencies should ensure that interpretive guides are used and referenced on all inspections.

  Interpretive guides should be made available to all stakeholders.

  Interpretive guides should be updated regularly as new or altered interpretative guidance is developed, for example, in response to experience with requirements or as new products or methods arise that raise questions about the applicability of, or options to comply with, a requirement.

- **Conducting team inspections.** Having multiple regulators on an inspection allows for consultation with peers to ensure that everyone agrees on whether something should be cited as a violation. Compliance determinations should be made independently and then the regulators can discuss any discrepancies afterwards.

- **Seeking guidance from other professionals.** Seeking guidance as soon as possible upon identifying an unusual situation or when there is a dispute about whether a violation exists is a very effective way to achieve interrater reliability. Regulators should be equipped with smartphones whenever possible for purposes of real-time consultation with supervisory staff, legal counsel, or program professionals.

- **Developing and modifying effective inspection tools.** Inspection tools such as measurement instruments and standardized forms should be developed and used such that they promote consistency without leading to inflexibility and rigidity that can arise when consistency is confused with uniformity.

- **Practicing peer reviews on a regular basis.** Regulatory oversight agencies should develop and conduct peer review exercises on a regular basis. Peer review exercises should include regulators, supervisors, and management staff, and should be conducted such that each participant's contribution of the discussion has equal weight. Peer review exercises may include:

  - Presenting scenarios where potential violations could be found to discuss how to measure compliance, which regulation should be cited, and how the description of the violation should be written;
  - Sharing actual descriptions of violations (with the regulator's name redacted) for constructive criticism; and
  - Performing "cross-region" inspections where one or more regulators from a given geographic location is paired with a regulator from another geographic region.

Interrater Reliability Assessment and Improvement

Regulatory oversight agencies should assess the strength and accuracy of staffs' interrater reliability and implement improvement strategies if weaknesses are found. The process for doing so includes, but is not necessarily limited to:

1) **Ensuring that all levels of agency personnel and stakeholders understand the tenets of interrater reliability.** It is essential that agency staff, providers, advocates, and the general public understand what interrater reliability is, why it is important, and its limitations.

2) **Conducting quantitative and qualitative analysis to measure interrater reliability strength.** Inspection data, provider feedback about the inspection experience, and other information must be collected and analyzed to identify interrater reliability strengths and weaknesses. While this is the most time consuming part of the process, it is essential in determining the interrater reliability.

3) **Evaluating the effectiveness of current practices.** If problems with interrater reliability are found—including consistent but inaccurate compliance measurement and determinations—the agency's current practices related to interrater reliability, if any, must be evaluated for effectiveness.

4) **Creating and modifying policy and instruments to improve consistency.** Once the scope and basis for weak or inaccurate interrater reliability are identified, the agency must create or modify interpretive guides, inspection practices, and work tools necessary for improvement.

5) **Providing robust staff training on new or revised policies and instruments.** Regulators must have a rich understanding of what has changed, why it has changed, and how to apply and use new policies and tools. . Trainings should include hands-on exercises to sharpen the regulator's skills. This should be repeated as necessary.

6) **Repeat the process.** Interrater reliability effectiveness and corrective actions should occur at least every three years to ensure ongoing program integrity.

About the National Administration for Regulatory Administration

NARA is a non-profit, professional association for human care regulators in adult care, child care, and child welfare who promotes the safety of children and adults through the study and development of best practices in the formulation, application and enforcement of licensing statutes and rules, producing educational materials and professional development resources on regulatory administration, and the provision of technical assistance to human care licensing agencies regarding the operation and management of the regulatory process.

For more information about NARA, please visit www.naralicensing.org.

For information about the professional services, including interrater reliability consulting services, contact NARA at admin@naralicensing.org.