Inter-Rater Reliability: Regaining Credibility with your Staff and Financial Officer While Meeting JCAHO Standards

Betty J. Noyes, RN, MA

*Journal of Nursing Administration*, Vol. 24, No. 9 (September), 1994, pp. 7-8

The Joint Commission on the Accreditation of Healthcare Organizations (JCAHO), in Standard NC 3.4.2., currently requires that our method for determining staffing has both face validity and inter-rater reliability. Furthermore, in LD.1.3, they require a plan to provide services in response to identified patient needs. In LD.2.1.4, the standards require that the leaders recommend a sufficient number of qualified persons consistent with the assessed needs of the population served and planned care/services. Every area must have process to determine the level of care required by patients being served in that specific area. The burden of proof that the required number and mix of staff members are provided will fall to patient care leadership.

In the past, creating a reasonable scientific basis for staffing methods brought "efficiency experts," later called "management engineers," to help the nursing department develop work load measurement instruments. These instruments helped nursing administrators enhance and preserve the credibility of their management decisions. This objective instrument theoretically would eliminate any biases in the staffing cycle.

Many came to believe that the indicator-based patient classification systems (PCS) was the best because it had the most discriminating design. It identified tasks that were labor sensitive and differentiated the nursing work load activity from the acuity of the patient illness. The PCS was believed to be reliable, objective, easy to understand, and controllable. Proponents also believed that PCS could be used to "flex" staff members accurately on a shift-by-shift basis.

Many chief nurse executives (CNEs) quickly learned that PCS was not a panacea for accurate staffing. The credibility of PCS slipped because we (CNEs) and the nursing staff learned that point values could be "tweaked" to reduce or expand staffing levels requirements. This was true for the projected numbers of nurses required to meet the demand levels and in skill mix. Skill mix was decided on arbitrarily. Patients were getting "more acute" because of shortening length of stay, and work loads were changing. Financial constraints were an ever-increasing reality.

The audiences of the PCS data - chief executives, chief financial officers, human resource directors, and nursing management and staff - were disappointed. The
audiences lost interest in hearing us defend budgets that were based on an unreliable instrument. Validity was destroyed every time a new program or "restructuring" occurred. Many organizations never made the effort to re-evaluate their PCS for new programs. Frequently, there was no money to rehire management engineers to collect new time/work load data.

Many CNEs questioned the effort spent reconstructing such a fallible "nursing" system. That has additional validity today because many organizations which have decentralized ancillary support services and a multiskilled work force do not have a PCS that measures the new task inventories. These redesigned care delivery systems require new methods to determine adequately the required staffing patterns. Decisions need to be made in instrument design and measurement methodology. This is a complex task because ancillary support services, such as physical therapy, respiratory therapy, phlebotomy, and EKG, have not had time/work load analysis done before, either in isolation or as part of the work load of a new, multiskilled, cross-trained worker.

Inter-rater reliability and face validity must be documented for which ever system is used to determine staffing requirements for the model of care in use. Validity ensures that the instrument measures the concept and the properties it is intended to measure. Face validity requires that the system appears to represent its intended purpose. Specifically, does it accurately reflect care requirements of patients in accordance with the practice patterns?

Reliability is concerned with consistency and repeatability. It questions whether the system produces consistent results when the measure is repeated on the same patient by different rates. Inter-rater reliability (IRR) is the measurement of two or more raters that classify the same patient. Inter-rater reliability does not measure the validity of a system. A system can be invalid and still have high IRR scores. These two diagnostic tools tell you only whether your system is healthy or moribund. Acceptable IRR must be achieved for each patient care division, unit, and type, and each frequently used indicator/criteria or prototype description. Inter-rater reliability goals are statistically satisfactory when IRR for individual geographic defined units or aggregated patient groups is in the 90 to 95 percentile range. Each indicator should have at least 80% agreement.

Sample size is another imperative. Each nurse involved in measuring IRR should obtain a 90% IRR score on a minimum of eight patients per quarter, representing a diversity of patient types. Each unit should measure a minimum of 15 to 20% of its unit census, never less than five patients per quarter. The final composite report needs to be for the entire patient care department.

When evaluating your PCS, it is helpful to measure IRR and face validity on each indicator or prototype description and for each participating staff nurse. This frequently will illuminate the source of any discrepancy and reason for low total IRR scores.

Approaches to measure IRR, although seemingly simplistic, are very labor intensive for a busy patient administration/staffing office. The importance of both measures is well worth the effort.
The approach to IRR is to designate monitors. These monitors must be selected carefully for their objectivity and clinical knowledge, then educated about the PCS. Monitoring staff members should be trained so that they can achieve 95% IRR scores among all patient types/classes before the IRR study begins.

The scheduling of the IRR study also requires orchestration. The monitor should classify the patient within 2 hours of the staff/unit level personnel to minimize the possibility of drastic shifts in the patient condition. Should a drastic shift occur within 2 hours, that patient's classification should be discarded. You are trying to measure IRR, not the differences caused by patient changes in condition.

The amount of time required for tabulation of these results is also significant. Many data points and correlations are required.

Inter-rater reliability results are important when the face validity discussions begin. Your experts in face validity are your staff, middle manager, and statistical data derived from other sources, including your IRR study. Face validity questions whether your instrument appears reasonable, complete, and representative of the nursing care standards and practices of your institution. Face validity can be achieved if clinicians and management periodically review all the components of the staffing system. They need to disassemble the parts and evaluate and discuss the definitions that are critical to any PCS validity, the methodology, and the IRR scores for each patient unit designation and each patient type. They must review that the used PCS instrument is reflecting current safe practice of patient care, that the current system responds to changes in the patient condition/work load, and that the resulting staffing recommendations appear reasonable and match patient care requirements and budgetary reality.

Results of these discussions frequently result in instrument and method redesign. Recommendations then are made by this face validity task force to patient care/hospital management for executive decision and ultimately, are included in the budgetary planning and scheduling/staffing guidelines.

The JCAHO requirement has given management the opportunity to revisit the validity of our PCS as a system that drives the fundamental decisions about the use of employees. It is our opportunity to reassure our current staff that the administration is accurately defining their present work load at the bedside. Our PCS must include all patient care providers, including a cross-trained, multiskilled worker as part of the patient care delivery model. No PCS should take precedence over on-site crisis management decisions regarding the need for staffing on a given unit at a given time. A PCS with good face validity and IRR is the first step in developing harmony and balance of human resource supply with the patient care work load demands of the bedside provider staff.

Reference

IL: Joint Commission on the Accreditation of Healthcare Organizations; 1993.