Comparison of the Canadian C-Spine Rule and NEXUS Decision Instrument in Evaluating Blunt Trauma Patients for Cervical Spine Injury

William R. Mower, MD, PhD
Jerome Hoffman, MD, MA

From the University of California–Los Angeles Emergency Medicine Center, Los Angeles, CA.

See related article, p. 507, and editorial, p. 518.

Concerned about missing potentially catastrophic neurologic injury, emergency physicians have typically made liberal use of radiographic imaging to evaluate blunt trauma patients for cervical spine injuries. This practice subjects large numbers of patients to imaging, with its associated cost, time expenditure, and radiation exposure, in order to detect injury in a small minority. Consequently, decision instruments that allow clinicians to safely reduce cervical spine imaging have the potential to be of significant value.

One existing instrument, the National Emergency X-Radiography Utilization Study (NEXUS) low-risk criteria, has been shown in prospective application to more than 34,000 patients to have a sensitivity of 99.6% for detecting clinically important cervical spine injury.1 However, in this issue of Annals, Canadian researchers, seeking to develop their own decision instrument, report the NEXUS instrument to have a sensitivity of less than 93% when retrospectively applied to their patient population.2 These results are inconsistent with the voluminous data collected during the development and validation of the NEXUS instrument and are in conflict with the large body of literature that investigated similar criteria before the performance of the NEXUS trial. Furthermore, the reported 7% miss rate with the NEXUS criteria is inconsistent with clinical experience and existing medical literature, which, except for the rare cases presented in the original NEXUS report, is virtually devoid of reports of missed injury.

The discrepancy between these 2 studies reflects, in part, a natural asymmetry between the processes of “validating” and “invalidating” a decision instrument. Validation studies are quite vulnerable to misclassification errors, and when such errors occur, an instrument can easily appear to have been “invalidated.” For example, failure to detect important clinical findings because of inadequate evaluations or the use of surrogate variables can cause high-risk patients to be misclassified as low risk. If injured high-risk patients are misclassified in this manner, the reported sensitivity and negative predictive value will decrease. In contrast, if uninjured patients who exhibit high-risk criteria are misclassified, the reported specificity may be falsely increased. Misclassification can also decrease the instrument’s interrater reliability, as reflected by measures such as the $\kappa$ statistic.

Compared with the performance of the NEXUS decision instrument documented in the NEXUS report, the Canadian study reports a large decrease in sensitivity, a small but real decrease in negative predictive value, and an increase in the measured specificity.1,2 This pattern suggests the presence of misclassification errors in the Canadian study, and likely results from the study’s retrospective methodology and use of surrogate variables. Consequently, the Canadian article tells us little about the true performance of the NEXUS instrument, but does serve as an important warning regarding the use of decision instruments in general. Clinicians who wish to use a given instrument must understand the definitions used by the instrument, and they must perform careful assessments in determining the classification of individual patients. Failure to use a decision instrument properly can produce inadequate and misleading assessments, can produce misclassification of risk status, and can have potentially devastating consequences.

Bearing these concerns in mind, clinicians should retain confidence in the reliability of the NEXUS cervical spine criteria. This instrument has already undergone validation in a large prospective study involving a wide range of institutions, clinical settings, and clini-
Comparing the Canadian C-Spine Rule and NEXUS

Decision instruments are based on the assessment of 5 criteria and can be used to determine which patients actually have cervical spine injury. Consequently, the principal benefit of both instruments lies in their ability to identify safely patients who do not require imaging. They are much less efficient in determining which patients actually have cervical spine injury.

Although these 2 instruments have substantial differences, they also share common strengths and weaknesses. Decision instruments are vulnerable to inadequate assessments and the use of surrogate variables, and the NEXUS and Canadian C-Spine Rule instruments are no exceptions. The application of either instrument requires clinicians to be familiar with the defining criteria and to conduct careful and deliberate assessments. Inappropriate application of either instrument can result in patient misclassification and serious medical error. It is also worth noting that these instruments have only been tested in blunt trauma patients during their acute presentations, and neither instrument has a demonstrated role in delayed assessment or for assessing patients who present with penetrating trauma.

These instruments exhibit statistically indistinguishable high sensitivity in identifying patients who have cervical spine injury, and both have similar excellent negative predictive value for excluding injury among patients identified as low risk—characteristics that are crucial in screening for potentially catastrophic injury. In contrast, neither instrument exhibits very high positive predictive value, and the vast majority of “non–low-risk” patients do not have cervical spine injury. Consequently, the principal benefit of both instruments lies in their ability to identify safely patients who do not require imaging. They are much less efficient in determining which patients actually have cervical spine injury.

Important differences between the 2 instruments center mainly around the criteria they use and in their ease of clinical implementation. The NEXUS instrument is based on the assessment of 5 criteria and can be applied to all acute blunt trauma patients. The Canadian C-Spine Rule is more complex and relies on a series of evaluations. It also has several inclusion criteria that limit its application in some patient groups, including children and pregnant women.

Differences between the 2 instruments could actually work in favor of astute clinicians who are familiar with both instruments. Although clinicians may prefer one instrument because it is easier to use or more applicable, patients who cannot be readily evaluated by this instrument may be amenable to evaluation by the other. It may also prove possible to combine the 2 in some instances, with the maintenance of high sensitivity and a useful increase in positive predictive value. Consequently, efforts to define a “best” instrument are misleading, unnecessary, and unproductive.

Although other issues have been mentioned as representing important differences between the 2 instruments, most are relatively unimportant. Concerns related to methodologic development of the 2 instruments are essentially meaningless. Recommendations for developing decision instruments do not guarantee the performance of the final instrument. The overall value of an instrument must be determined by appropriate testing in a validation study that has sufficient size to reach statistically meaningful conclusions and that is conducted in environments (eg, patients, clinicians, institutions) similar to those in which the instrument will be used. Derivation studies and interrater analyses are useful means for identifying potentially important criteria, but the actual value of any criterion or instrument and its reliability in the clinical setting can only be assessed through validation.

Differences between the NEXUS and Canadian measurements of interrater reliability undoubtedly reflect misclassification errors in the Canadian study and likely derive from their use of surrogate variables. However, it is important to note that interrater reliability in itself is not a sufficient reason to include or exclude an individual criterion from consideration in a decision instrument. Criteria with overall poor interrater reliability may nevertheless be very sensitive and even reliable indicators of disease, whereas poor overall performance may reflect difficulties in assessing the criterion among unafflicted patients. Such criteria may be particularly valuable in constructing instruments with high sensitivity, as required by NEXUS and the Canadian C-Spine Rule, although their inclusion may decrease specificity and positive predictive value as a result of frequent misclassification of patients without injury. Measures of interrater variability serve as surrogate markers for predicting how well a given clinical finding might perform when applied by a large number of clinicians.
observers; the true utility of any criterion can only be assessed through an appropriate validation study.

The difference in the measured specificity of the 2 instruments is also not particularly meaningful. Specificity describes the proportion of uninjured patients who are properly classified by each instrument. Because the 2 studies used different methodologies, the measured specificities are expected to be different. The Canadian C-Spine study enrolled all patients who had sustained neck trauma, including a sizeable proportion who did not undergo radiographic evaluation. In contrast, the NEXUS study enrolled only patients who underwent radiographic imaging and specifically excluded patients who did not. Patients who did not undergo imaging represent a sizeable proportion of the uninjured, and many of these individuals do not exhibit risk criteria from either rule. By excluding these patients, the NEXUS study decreased the number of potentially uninjured patients that were evaluated and consequently found a lower specificity than would have been found if they had been included. The Canadian C-Spine study included these cases in their calculations and consequently found a higher specificity.

Indirect evidence suggests that the 2 instruments actually have very similar specificities among equivalent cohorts. For example, the expected impact on current radiographic imaging is very similar for each of the instruments, even though ordering patterns in the NEXUS study were more efficient than those in the Canadian C-Spine study (achieving almost twice as high a fracture rate), and despite the fact that film ordering in NEXUS had already decreased substantially after publication of the derivation study.6

Assuming the Canadian C-Spine Rule is validated in a wide variety of clinical settings, future work may reveal additional differences between the 2 instruments, but it is unlikely that such work will greatly affect the benefits that derive from having 2 well-validated decision instruments to guide the selective cervical spine imaging of blunt trauma patients. In this regard, the development and validation of the NEXUS criteria and the Canadian C-Spine Rule represent an embarrassment of riches for emergency physicians, who will have the luxury of choosing to use either or both of these instruments, depending on their clinical setting, ease and applicability, and individual practice styles.

The authors report this study did not receive any outside funding or support.

Reprints not available from the authors.