agents. Thalidomide, once infamous, has redeemed itself as a potent immunomodulatory agent that induces responses even in patients with relapsed or therapy-resistant myeloma. Thalidomide therapy is a viable initial alternative to chemotherapy when combined with dexamethasone. Bortezomib, a proteasome inhibitor, has substantial single-agent activity against myeloma. Studies are under way that incorporate treatment with this drug early in the course of the disease, as are trials of bortezomib in combination with other agents. How such developments will alter therapy for myeloma remains to be seen, but these new chapters in the story have greatly expanded the options for the treatment of patients with this disease. The year 2004 will be an encouraging time for patients with myeloma, but the best is yet to come. Stay tuned.

From the Bone Marrow and Stem Cell Transplant Program, Abramson Cancer Center, University of Pennsylvania, Philadelphia.


Choosing between Clinical Prediction Rules

Donald M. Yealy, M.D., and Thomas E. Auble, Ph.D.

Clinicians currently order imaging studies for most patients with blunt head or neck trauma if there is even a remote possibility of cervical-spine injury, in order to avoid missing a potentially disabling fracture, dislocation, or ligamentous injury. The result of this practice is that 96 percent or more of cervical-spine radiographs reveal no clinically important injuries. A decision rule, or clinical prediction rule, that could reliably identify patients with trauma who are at very low risk for cervical anatomical disruption could help reduce unnecessary radiographic testing, improve efficiency, and decrease health care expenditures that accompany testing.

In this issue of the Journal, Stiell et al. compare the performance of two evidence-based clinical decision rules designed to provide this aid to clinicians: the Canadian C-Spine (cervical-spine) Rule (CCR) and the National Emergency X-Radiography Utilization Study (NEXUS) Low-Risk Criteria (NLC). The questions we address here are simple: Is the CCR valid, and is one of these rules better than the other? What evidence for each rule existed before the current study? The CCR was derived from enrolling patients with cervical trauma. The CCR uses a three-step process to identify patients for whom radiography is unnecessary; the variables assessed are risk factors for serious injury (including the patient’s age and the mechanism of injury) and the findings on physical examination, including the ability of patients to rotate their necks actively. In the derivation study, the CCR had 100 percent sensitivity and 42.5 percent specificity for identifying important cervical-spine injuries. The NLC incorporated five criteria that smaller studies suggested could identify patients who were at very low risk for injury and for whom imaging might be unnecessary. In a prior prospective validation study involving more than 34,000 patients un-
ndergoing cervical-spine radiography, the NLC had 99.6 percent sensitivity and 12.9 percent specificity for identifying serious injury. The higher specificity of the CCR, as compared with the NLC, seemed to be largely attributable to the inclusion of patients who did not have cervical-spine radiography (such patients were excluded from the NEXUS cohort). Thus, available data suggested that both rules were reasonable and safe, although different.

In the current study, Stiell and colleagues compared the performance of these two rules in 8283 patients with head or neck trauma in nine Canadian hospitals. In this study population, the CCR had nearly perfect sensitivity (99.4 percent) and moderate specificity (45.1 percent) for detecting clinically important cervical-spine injuries — findings that are consistent with those in the earlier derivation study. In contrast, the sensitivity of the NLC in this study dropped to 90.7 percent, which is equivalent to missing serious injuries in nearly 10 percent of patients classified as being at “low risk.” Furthermore, the NLC had lower specificity, suggesting that the use of this rule would confer less ability to reduce unnecessary testing. The authors conclude, “These findings raise questions about the safety and efficiency of applying the NLC in clinical practice.”

Although the present data validate the performance of the CCR, the study design has several limitations that call into question any conclusion about the superiority of one rule over the other. The NLC was tested only in patients for whom cervical radiographs were ordered, and the study population included patients 16 years of age and under and those with an altered level of consciousness. All of these patient-eligibility criteria differ from those used in the CCR study, making comparisons of the two rules in a cohort that meets CCR criteria problematic. For example, excluding patients with an altered level of consciousness effectively negates the corresponding NEXUS criterion.

The choice of study sites may also have affected the performance of these rules in the current study. Stiell et al. enrolled patients at the same sites where the CCR was derived. Even if many of the physicians who enrolled patients and provided data did not participate in the earlier derivation study, the potential local and regional familiarity with the CCR might have improved the performance of this rule relative to the performance of the NLC. The finding that clinicians documented the presence of the CCR risk factor “paresthesias in extremities” but failed to document the similar NEXUS criterion “focal neurological deficit” in three patients with injuries (all counted as “missed” by the NLC) supports this concern.

The side-by-side comparison of the rules is also a source of potential bias. The authors attempted to address this problem in several ways. First, each rule was delineated on a data form (with the wording and definitions for the NLC approved by NEXUS investigators). However, these printed rules were not used to assess the actual variables; instead, a structured set of questions was used that may not have defined the variables as intended by the authors of the rule. To attempt to address this issue, the investigators also asked physicians about their degree of comfort with both rules and performed a post hoc review to validate the data independently. However, these actions cannot replace testing in settings where physicians have equal pretrial experience with both rules or have been required to use each as intended.

Examination of the transportability of a rule implies that the same definitions of predictors and outcomes used in the first setting are used in the new setting. In the current study, the authors used the same definitions that had been used to derive the CCR; the same standard was not applied to the NLC. Instead, the authors added “clarifying” definitions to each NEXUS risk factor. These post hoc changes are seemingly at odds with the explicit decision of the NEXUS investigators not to include such clarifications in the original rule. Likewise, the definition of “intoxication” was narrowed, potentially reducing the sensitivity of the NLC. The poor performance of the NLC in the retrospective Canadian assessment of the NLC suggests that these methodologic problems may have existed in both attempts to compare the performance of the NLC and CCR.

Another possible reason that the performance of the rules differed is the content of each rule. For example, the CCR involves assessment of neck movement to determine the need for imaging, whereas the NLC is silent on this feature. At first glance, some physicians may be uncomfortable with this neck-movement recommendation; even in the current study, clinicians did not test this in 10 percent of the patients in spite of the request. However, to reach this step in the CCR, antecedent criteria must all be suggestive of the absence of injury. The current and previous CCR data, coupled with practical experience, support the safety of this criterion.
current study, the apparently greater sensitivity of the CCR than of the NLC can be attributed to its component variables and overall efficacy or whether it simply reflects methodologic issues.

Decision rules that are data-driven have the potential to optimize care by limiting unnecessary variation while maintaining or improving outcomes. The current study makes clear the potential for the CCR to achieve this goal among patients with cervical trauma, although more work is needed.

The last step before widespread application of a clinical prediction rule is to test its effect on patient care.\(^4,5\) Although both the CCR and NLC have been independently validated,\(^1,2\) neither has been examined in a controlled implementation trial. Outside validation is needed, since performance drop-off is common when a rule is used in a new setting;\(^6,7\) likely reasons include differences in physicians’ training and experience, implementation tools, and target populations between the initial research setting and the new setting. Although the current study calls into question the reproducibility of the original NEXUS validation study, we believe that it is premature to abandon either rule, but that it is also premature to embrace either rule without additional research and experience. Ultimately, the true value of a rule depends on its performance in routine practice.

We are indebted to Margaret Hsieh, M.D., M.S., for her thoughtful assistance in preparing this work.

From the Department of Emergency Medicine, University of Pittsburgh, Pittsburgh.


Copyright © 2003 Massachusetts Medical Society.