Measuring Acceptability of Clinical Decision Rules: Validation of the Ottawa Acceptability of Decision Rules Instrument (OADRI) in Four Countries

Jamie C. Brehaut, Ian D. Graham, Timothy J. Wood, Monica Taljaard, Debra Eagles, Alison Lott, Catherine Clement, Anne-Maree Kelly, Suzanne Mason and Ian G. Stiell

Med Decis Making 2010 30: 398 originally published online 30 December 2009
DOI: 10.1177/0272989X09344747

Measuring Acceptability of Clinical Decision Rules: Validation of the Ottawa Acceptability of Decision Rules Instrument (OADRI) in Four Countries

Jamie C. Brehaut, PhD, Ian D. Graham, PhD, Timothy J. Wood, PhD, Monica Taljaard, PhD, Debra Eagles, MD, Alison Lott, BSc, Catherine Clement, RN, Anne-Maree Kelly, MD, MClinEd, FACEM, Suzanne Mason, MBBS, FRCS, FFAEM, MD, Arthur Kellerman, MC, MPH, Ian G. Stiell, MD, MSc, FRCPC

Background. *Clinical decision rules can benefit clinicians, patients, and health systems, but they involve considerable up-front development costs and must be acceptable to the target audience. No existing instrument measures the acceptability of a rule. The current study validated such an instrument.* ***Methods.*** *The authors administered the Ottawa Acceptability of Decision Rules Instrument (OADRI) via postal survey to emergency physicians from 4 regions (Australasia, Canada, United Kingdom, and United States), in the context of 2 recently developed rules, the Canadian C-Spine Rule (C-Spine) and the Canadian CT Head Rule (CT-Head). Construct validity of the 12-item instrument was evaluated by hypothesis testing.* ***Results.*** *As predicted by a priori hypotheses, OADRI scores were 1) higher among rule users than nonusers, 2) higher among those using the rule ''all of the time'' v. ''most of the time'' v. ''some of the time,'' and 3) higher among rule nonusers who would consider using a rule v. those who would not. We also examined explicit reasons given by respondents who said they would not use these rules. Items in the OADRI accounted for 85.5% (C- Spine) and 90.2% (CT-Head) of the reasons given for not considering a rule acceptable.* ***Conclusions.*** *The OADRI is a simple, 12-item instrument that evaluates rule acceptability among clinicians. Potential uses include comparing multiple ''proto-rules'' during development, examining acceptability of a rule to a new audience prior to implementation, indicating barriers to rule use addressable by knowledge translation interventions, and potentially serving as a proxy measure for future rule use.* ***Key words:*** *acceptability; clinical decision rules; validation; survey.* ***(Med Decis Making 2010; 30: 398–408)***

Received 3 October 2008 from Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, ON, Canada (JCB, IDG, MT, DE, AL, CC, IGS); Department of Epidemiology and Community Medicine, University of Ottawa, Ottawa, ON, Canada (JCB, IDG, MT, IGS); School of Nursing, Faculty of Health Sciences, University of Ottawa, Ottawa, ON, Canada (IDG); Knowledge Translation Portfolio, Canadian Institutes of Health Research, Ottawa, ON, Canada (IDG); Medical Council of Canada, Ottawa, ON, Canada (TJW); Department of Emergency Medicine, University of Ottawa, ON, Canada (DE, IGS); Joseph Epstein Centre for Emergency Medicine Research at Western Health and the University of Melbourne, Melbourne, Australia (A-MK); School of Health and Related Research, University of Sheffield, Sheffield, UK (SM); and Department of Emergency Medicine, Emory University, Atlanta, GA (AK). Financial support for this study was provided by a grant from the Canadian Institutes of Health Research (FRN# MOP-86709). The funding agreement ensured the authors' independence in designing the study, interpreting the data, and writing and publishing the report. The following author is currently on secondment to CIHR: IDG. Revision accepted for publication 1 July 2009.

Clinical decision rules (rules) are algorithmic decision tools designed to help clinicians predict an important clinical outcome with known sensitivity and specificity by considering a small number of highly valid indicators. Some rules are widely used,[1–4] as growing numbers of clinicians

To view Web Appendices 1–3, the reader should refer to http://mdm.sagepub.com/supplemental/. Address correspondence to Jamie C. Brehaut, PhD, Scientist, Ottawa Hospital Research Institute, Associate Director, Bell Patient Decision Support Laboratory, Career Scientist, Ontario Ministry of Health and Long Term Care, Ottawa Health Research Institute, Clinical Epidemiology Program, Ottawa Hospital, Civic Campus, ASB 2-004; Box 693, 1053 Carling Avenue, Ottawa, ON K1Y 4E9, Canada; telephone: (613)798-5555 ext. 13432; fax: (613) 761-5402; e-mail: jbrehaut@ohri.ca.

DOI: 10.1177/0272989X09344747

realize that clinical judgment alone can be inconsistent and flawed.[5–7] Use of high quality rules has been shown to benefit patients (reduced waiting times, reduced unnecessary procedures), physicians (easier, more evidence-based decisions), health care systems (improved efficiency), and payers (reducing costs associated with unnecessary procedures).[1,8–16]

High quality rules require a rigorous and resource-intensive development process involving several stages and thousands of patients. Standards for the development of rules have been outlined elsewhere.[17–19] Typically, they include background work determining the need for a rule, derivation involving data collection on a wide variety of correlates of the outcome, prospective validation and refinement of the rule across multiple sites,[16,20,21] and large-scale implementation trials[11,13,22,23] and impact analyses.[24,25] Thus, although rules have many benefits after development and implementation, the initial costs are high.

Although evidence on the extent of rule use is limited,[17] the evidence that does exist suggests considerable variability in the extent to which rules will be adopted into practice. Some (e.g., the Ottawa Ankle Rules)[11]have been shown to be in wide use internationally, whereas other similar rules are used much less widely.[4] Given the resources involved in their development, an instrument that could measure whether a rule will be acceptable to clinicians would be extremely useful. If acceptable, the rule might warrant expending the resources required to complete its development. If not, the rule could be altered, abandoned, or subjected to further research; this approach has been used with clinical guidelines.[26,27]

To serve this need, we developed the Ottawa Acceptability of Decision Rules Instrument (OADRI), a 12-item instrument designed to measure the acceptability of a clinical decision rule (Web Appendix 1). This report describes the development, pilot testing, and initial validation of this instrument in the context of surveys about 2 rules: the Canadian C-Spine Rule (C-Spine)[14,16] and the Canadian CT Head Rule (CT-Head).[15,21]

VALIDATION

When there is no generally accepted measure to serve as the ''gold standard,'' construct validity of an instrument (i.e., whether we are truly measuring rule acceptability with our instrument) should be evaluated through hypothesis testing and discriminant validity.[28] Scores should be correlated with factors related to the target construct and should discriminate between subgroups that would be expected to differ on the target construct. We generated 6 hypotheses designed to test the construct validity of the OADRI:

1) Those who report using a rule should score higher on the OADRI than those who report not using the rule.
2) Among those who report using the rule, OADRI scores should be higher among those that report using the rule consistently.
3) Among those who do not currently use the rule, those who say they would consider using it in the future should score higher on the OADRI than those who would not consider using it.
4) OADRI scores should be higher in the region where the rule was developed.
5) OADRI scores should be higher for more established rules (i.e., the C- Spine) than for less well-established rules (CT-Head).
6) Reasons given for not using a rule should be accounted for by 1 or more items in the OADRI.

More details on these hypotheses are described in Web Appendix 2.

DEVELOPMENT

Many of the OADRI items have been developed over years of research involving surveys,[1,3,4,29] interviews with emergency physicians,[1,2] and studies of guideline development.[30–32] Over time, themes have emerged about the reasons physicians give for why they will or will not adopt rules into practice. Three of the authors (JB, IS, IG) initially assembled to evaluate a variety of items that addressed known aspects of rule acceptability. This process was informed by the Ottawa Model of Research Use,[33,34] a framework that describes various barriers and facilitators related to whether a research innovation will be adopted into practice and groups them into 3 large categories (aspects of the innovation, decision maker, and environment). An iterative process of editing and informal pilot testing resulted in a number of items being dropped or modified, leaving 15 items for pilot testing.

The 15-item proto-instrument was pilot tested on 12 practicing emergency physicians from a single tertiary care hospital. All of the pilot respondents were familiar with rules and their development, as this hospital has been at the center of the development of a number of rules. Pilot respondents were told about the purpose of the study and asked to complete the

15-item instrument twice: once in the context of a rule they used in their practice (their choice), and once in the context of another rule they were aware of but did not use in their practice. Open-ended questions/comments about the instrument were elicited. On the basis of their responses, several items were reworded and 3 were eliminated. One was dropped because of lack of variability in the responses, and 2 were dropped because they were highly correlated (>0.85) with another item in the instrument. Pilot results showed that the instrument was easy to complete, taking 2 to 3 min on average.

Evaluation

The current work was conducted as part of a larger survey studying perceptions about a variety of clinical decision rules among emergency physicians in 4 regions. Between the months of July 2005 and January 2006, 4 similar (but not identical) surveys were conducted of random samples of emergency medicine physicians drawn from emergency medicine associations in Australasia (AUS), Canada (CAN), the United Kingdom (UK), and the United States (US). Details about the survey protocols are summarized in Web Appendix 3. In terms of the items entering into the current analyses, the 4 surveys differed very little. Surveys were typically 7 pages long, and took between 15 and 20 min to complete. The 1st 2 pages of the survey described and asked questions about the C-Spine. Pages 3 and 4 addressed the CT-Head. Questions prior to the administration of the OADRI included 1) whether the clinician had been aware of the rule prior to the survey, 2) how often they apply the rule in their practice, and 3) if they currently do not use the rule, whether they would consider using it in the future, and 4) if not, why not.

Scoring and Analysis of the OADRI

Respondents were asked to indicate their level of agreement with each of the 12 statements on a 6-point scale ranging from 1 (Strongly Disagree) to 6 (Strongly Agree), or indicating ''No Opinion/Don't know'' (Web Appendix 1). The 1st 7 items were phrased such that higher numbers indicated greater acceptability, and the opposite was true of the last 5. We chose this strategy as a compromise between attempting to avoid yea-saying bias[28] and making the instrument as easy to complete as possible. Respondents completed the instrument once regarding the C-Spine and a 2nd time regarding the CT-Head. Order of presentation of the 2 instruments was not randomized to ensure parallel construction with other parts of the survey. If a respondent indicated he or she was not familiar with a rule prior to the survey, they were asked to review the rule (presented with the survey) and answer the 12 questions as if they were considering using the rule (the rule would be easy to use, etc.).

The final score consisted of the mean of all 12 items (recoded where necessary), resulting in a score that ranged from 0 to 6. Skipped items were excluded from the combined score. The mean of the remaining items served as the instrument score. Respondents completing less than 8 of the 12 items were counted as not having completed the instrument. Items for which ''No opinion/Don't know'' was selected were coded as the middle of the scale. We decided on this approach, rather than counting them as missing, because we felt both of these responses corresponded to equipoise on our scale, and because counting them as missing would have meant that the newer CT-Head rule would often have been assessed on fewer than 12 items. That said, we also conducted reanalyses for all of our major hypotheses, coding no opinion/don't know responses as missing, and found that no conclusions would have changed had we decided differently.

Analysis of the results began with descriptive analyses to determine the rate of uncompleted instruments, the rate of missing items among otherwise completed instruments, the mean and range of item scores, and interitem correlations. Internal consistency was measured with Cronbach's α, whereas item-total correlations showed the contribution of each item to the overall consistency of the instrument. These descriptive analyses were conducted on the overall data set, as well as separately for each of the 4 surveys and 2 instruments per survey.

More detailed analysis consisted of analyses of variance (ANOVA), with instrument scores as the outcome, to examine how different effects interact with Region in affecting instrument scores. For example, we conducted a 2 (C-Spine, CT Head) $\times$ 4 (Aus, Can, UK, US) within and between subjects ANOVA to examine the independent effects of Rule and Region on instrument scores, as well as their interaction (hypothesis 4). Similar analyses looked at Rule Use (Use, Don't Use) $\times$ Region (hypothesis 1), Consistent Use (use rule ''Sometimes,'' ''Most of the time,'' ''Always'') $\times$ Region (hypothesis 2), and Consider Use (among nonusers: Would consider, Would not consider use) $\times$ Region (hypothesis 3). For these latter 3 analyses, we examined effects for both C-Spine and CT-Head separately, to assess replicability. Post hoc

t tests (with Bonferroni adjusted significance levels) allowed us to evaluate specific differences between regions. Where appropriate, area under the ROC curve (the *c*-statistic) describing the ability of instrument scores to discriminate between subgroups was computed.

To determine if our instrument had adequate construct coverage, we asked all respondents who said they would not consider using a rule to indicate, via an open-ended question, what their reasons for nonuse were. We compiled these comments and examined whether they could reliably be associated with 1 or more of the instrument items. Two of the authors (JB and AL) rated each comment on 1) which instrument item(s) addressed the comment, 2) whether the comment was not covered by an instrument item, or 3) whether the comment was too unclear to code. Percent agreement and Cohen's κ were computed between 2 raters on whether reasons given for not using a rule corresponded to 1 or more instrument items.

Survey Administration

Within the constraints of the various national organizations involved, study protocol and mailouts followed the recommendations of Dillman's tailored design method.[35] No incentives were provided. Correspondence was addressed from IS (Canada), and from local, nationally prominent investigators: SM (UK), A-MK (AUS), and AK (US). Any nonphysician respondents were excluded from analysis. Details of the various survey protocols are presented in Web Appendix 3. A total of 1297 respondents returned a survey; country-specific response rates ranged from 44.5% to 69%.

RESULTS

Internal Consistency

Across all 4 surveys, the number of respondents who did not complete the instrument (defined as completing at least 8 of the 12 items) was very low (C-Spine: 17 of 1297, or 1.3%; CT-Head: 54 of 1297, or 4.2%). Table 1 describes the mean item scores, the item-total correlations, and the internal consistency (Cronbach's α) for each of the 4 surveys, and each rule. The range of individual item scores covered the entire range for all 12 items; overall percentage of items coded as ''No Opinion/Don't Know'' was 8.1% for C-Spine and 12.9% for CT-Head. Mean item scores tended to be above a score of 3 (the middle of the scale), indicating overall acceptability of the 2 rules (i.e., higher numbers indicate greater acceptability). Internal consistency was quite high across all groups, ranging from 0.79 to 0.85 for the C-Spine and from 0.80 to 0.86 for the CT-Head. Item-total correlations ranged from 0.20 to 0.72 for C-Spine and from 0.09 to 0.75 for CT-Head.

Hypothesis 1: Rule Use

Table 2 reports mean instrument scores related to the hypothesis that scores should be higher among those who say they use a rule than those who do not. Across all 4 regions, ANOVA showed that scores were significantly higher for those who reported using the rule than for those who did not ($F_{1,1242} = 273.6$, $P < 0.001$). A significant Use × Region interaction ($F_{3,1242} = 6.7$, $P < 0.001$) showed that the magnitude of this difference was not consistent across the 4 regions; the difference was smaller, but still statistically significant, for the US survey (post hoc simple effects analysis, $P < 0.001$). Similar results are shown in Table 2 for CT-Head acceptability scores. Across all 4 regions, scores were significantly higher for those who reported using the rule than for those who did not ($F_{1,1149} = 241.7$, $P < 0.001$)]. A nonsignificant Use × Region interaction ($F_{3,1124} = 0.47$, $P > 0.10$) showed that the magnitude of this difference was consistent across all 4 regions. *c*-Statistic describing the area under the ROC curve was 0.84 (C-Spine) and 0.79 (CT-Head), indicating good discrimination between users and nonusers for both rules.

Hypothesis 2: Consistent Use

Table 3 reports mean instrument scores related to the hypothesis that, among those who report using the C-Spine, scores should be higher among those who say they use it ''always'' than those who say they use it ''most of the time'' or only ''sometimes.'' ANOVA showed a main effect of Consistent Use ($F_{2,938} = 185.5$, $P < 0.001$). Post hoc multiple comparisons showed that those reporting using the rule ''sometimes'' scored significantly lower than those saying they use it ''most of the time'' ($P < 0.001$), who in turn scored significantly lower than those who reported using the rule ''always'' ($P < 0.001$). A nonsignificant Consistent Use × Region interaction ($F_{6,938} = 1.5$, $P > 0.10$) showed that this pattern held true for all 4 regions. Analysis of ratings of the CT-Head produced similar results. The main effect of Consistent Use was highly significant ($F_{2,669} = 92.1$,

Table 1 Item Means, Item-Total Correlations, and Internal Consistency (Cronbach's α) of the OADRI across 4 Countries and 2 Clinical Decision Rules, for Respondents Who Completed All 12 Items

	Canadian C-Spine Rule					Canadian CT Head Rule				
	AUS	**CAN**	**UK**	**US**	**Total**	**AUS**	**CAN**	**UK**	**US**	**Total**
n	398	311	271	221	1201	380	308	257	217	1162
Cronbach's α	0.85	0.82	0.79	0.78	0.83	0.86	0.85	0.80	0.81	0.84
Means (item-total correlations)										
1. Easy to use	4.30 (0.67)	4.84 (0.57)	5.13 (0.45)	4.82 (0.59)	4.72 (0.60)	4.73 (0.62)	5.05 (0.55)	4.96 (0.47)	4.78 (0.60)	4.88 (0.56)
2. Easy to remember	3.39 (0.59)	3.84 (0.46)	4.23 (0.43)	4.11 (0.48)	3.83 (0.50)	4.14 (0.50)	4.47 (0.42)	4.39 (0.36)	4.37 (0.54)	4.33 (0.45)
3. Useful in my practice	4.43 (0.72)	5.34 (0.59)	5.17 (0.66)	4.86 (0.63)	4.91 (0.68)	4.43 (0.74)	5.11 (0.66)	4.31 (0.67)	4.53 (0.67)	4.60 (0.70)
4. Wording is clear and unambiguous	4.36 (0.50)	4.76 (0.41)	4.80 (0.32)	4.62 (0.45)	4.61 (0.45)	4.47 (0.50)	4.92 (0.50)	4.62 (0.33)	4.47 (0.45)	4.62 (0.46)
5. My colleagues support use of the rule	3.95 (0.69)	4.53 (0.44)	4.51 (0.53)	3.59 (0.56)	4.16 (0.60)	3.75 (0.68)	4.37 (0.50)	3.47 (0.55)	3.36 (0.56)	3.78 (0.61)
6. Patients benefit from use of the rule	4.73 (0.68)	5.18 (0.59)	4.97 (0.67)	4.72 (0.61)	4.90 (0.65)	4.38 (0.72)	4.94 (0.70)	4.28 (0.71)	4.42 (0.63)	4.51 (0.71)
7. Results in improved use of resources	4.64 (0.55)	5.26 (0.55)	4.80 (0.57)	5.05 (0.54)	4.91 (0.55)	4.38 (0.59)	4.88 (0.61)	3.65 (0.55)	4.88 (0.51)	4.44 (0.56)
8. Would increase the chance of lawsuits (reversed)	4.42 (0.22)	4.61 (0.39)	4.56 (0.36)	3.43 (0.20)	4.32 (0.28)	4.23 (0.46)	4.39 (0.47)	4.30 (0.42)	3.17 (0.35)	4.09 (0.42)
9. Evidence supporting the rule is flawed (reversed)	4.40 (0.38)	4.80 (0.42)	4.29 (0.38)	4.01 (0.39)	4.42 (0.40)	3.97 (0.49)	4.41 (0.47)	3.95 (0.48)	3.65 (0.45)	4.02 (0.50)
10. Using another rule or similar strategy (reversed)	2.20 (0.46)	4.06 (0.52)	3.43 (0.41)	2.12 (0.27)	2.96 (0.47)	2.71 (0.43)	4.44 (0.42)	1.66 (0.28)	2.59 (0.09)	2.91 (0.39)
11. Does not account for important clinical cue (reversed)	3.52 (0.33)	4.25 (0.49)	3.68 (0.35)	4.04 (0.27)	3.84 (0.37)	3.49 (0.47)	4.19 (0.52)	3.51 (0.33)	3.43 (0.43)	3.67 (0.47)
12. Environment I work in makes it difficult to use (reversed)	4.35 (0.42)	4.87 (0.46)	4.82 (0.30)	4.10 (0.30)	4.55 (0.40)	4.17 (0.35)	4.60 (0.52)	3.77 (0.34)	3.69 (0.45)	4.11 (0.47)

Note: OADRI, Ottawa Acceptability of Decision Rules Instrument; AUS, Australasia; CAN, Canada; UK, United Kingdom; US, United States.

Table 2 Hypothesis 1—Mean (SE) Instrument Scores across Subsets of Respondents Who Report Using/Not Using the Rules

C-Spine	AUS (*n* = 278)	CAN (*n* = 306)	UK (*n* = 229)	US (*n* = 137)	Total
Don't Use Rule	3.18	3.69	3.67	3.73	3.57
	(0.07)	(0.17)	(0.11)	(0.09)	(0.06)
Use Rule	4.49	4.77	4.75	4.42	4.61
	(0.05)	(0.05)	(0.05)	(0.07)	(0.03)
CT-Head	**AUS (*n* = 389)**	**CAN (*n* = 315)**	**UK (*n* = 256)**	**US (*n* = 197)**	**Total**
Don't Use Rule	3.57	3.94	3.45	3.77	3.68
	(0.06)	(0.12)	(0.07)	(0.08)	(0.04)
Use Rule	4.44	4.83	4.40	4.47	4.54
	(0.06)	(0.05)	(0.08)	(0.10)	(0.04)

Note: C-Spine, Canadian C-Spine Rule; AUS, Australasia; CAN, Canada; UK, United Kingdom; US, United States; CT-Head, Canadian CT Head Rule.

Table 3 Hypothesis 2—Mean Instrument Scores (SE) across Subsets of Rule Users Who Report Using the Rules Sometimes, Most of the Time, or Always

C-Spine	AUS (*n* = 278)	CAN (*n* = 306)	UK (*n* = 229)	US (*n* = 137)	Total
Sometimes	3.87	3.82	4.07	3.86	3.91
	(0.07)	(0.09)	(0.10)	(0.09)	(0.05)
Most of the time	4.61	4.88	4.80	4.65	4.73
	(0.06)	(0.05)	(0.06)	(0.08)	(0.03)
Always	5.03	5.25	5.12	5.20	5.15
	(0.08)	(0.07)	(0.09)	(0.16)	(0.05)
CT-Head	**AUS (*n* = 221)**	**CAN (*n* = 263)**	**UK (*n* = 122)**	**US (*n* = 75)**	**Total**
Sometimes	3.91	4.15	4.07	4.11	4.06
	(0.07)	(0.08)	(0.09)	(0.12)	(0.05)
Most of the time	4.73	4.97	4.57	4.63	4.72
	(0.07)	(0.06)	(0.11)	(0.12)	(0.05)
Always	4.99	5.39	4.90	5.44	5.18
	(0.13)	(0.09)	(0.14)	(0.25)	(0.08)

Note: C-Spine, Canadian C-Spine Rule; AUS, Australasia; CAN, Canada; UK, United Kingdom; US, United States; CT-Head, Canadian CT Head Rule.

$P < 0.001$), and post hoc multiple comparisons showed significant differences between the ''sometimes'' group and the ''most of the time'' group ($P < 0.001$), and between the ''most of the time'' group and the ''always'' group ($P < 0.001$). Again, a nonsignificant Consistent Use × Region interaction ($F_{6,669} = 1.6$, $0 > 0.10$) shows that this pattern held true for all 4 regions.

Hypothesis 3: Nonusers, Consider Use

Table 4 reports mean instrument scores related to the hypothesis that, among those who report not using the C-Spine, scores should be higher among those who say they would ''consider'' using it than those who would ''not consider'' it. ANOVA showed a main effect of Consider Use ($F_{1,432} = 89.2$, $P < 0.001$) showing that instrument scores were significantly higher for those who would consider using the C-Spine than those who would not consider using it. A nonsignificant Consider Use × Region interaction ($F_{3,432} = 0.3$, $P > 0.10$) shows that this pattern was consistent across all 4 regions. Similar results were obtained for the CT-Head. The main effect of Consider Use was highly significant ($F_{1,652} = 99.2$, $P < 0.001$) showing that instrument scores were higher for those who would consider using the CT-Head than those who would not consider using it. The significant Consider Use × Region interaction ($F_{3,637} = 3.98$, $P = 0.008$) showed that the magnitude of this effect was not consistent across all regions. The difference in instrument scores between

Table 4 Hypothesis 3—Mean Instrument Scores (SE) across Subsets of Rule Nonusers Who Report They Would/Would Not Consider Using the Rule in the Future

C-Spine	AUS (n = 162)	CAN (n = 58)	UK (n = 85)	US (n = 135)	Total
Would not use	2.98	3.33	3.28	3.18	3.19
	(0.08)	(0.20)	(0.17)	(0.12)	(0.07)
Would consider use	3.90	4.08	4.04	4.08	4.03
	(0.09)	(0.12)	(0.10)	(0.08)	(0.05)
CT-Head	**AUS (n = 216)**	**CAN (n = 94)**	**UK (n = 171)**	**US (n = 179)**	**Total**
Would not use	3.00	3.58	3.40	3.00	3.25
	(0.09)	(0.21)	(0.07)	(0.12)	(0.07)
Would consider use	4.01	4.14	3.93	4.04	4.03
	(0.07)	(0.09)	(0.10)	(0.07)	(0.04)

Note: C-Spine, Canadian C-Spine Rule; AUS, Australasia; CAN, Canada; UK, United Kingdom; US, United States; CT-Head, Canadian CT Head Rule.

those who would and would not consider using the CT-Head was greater in the AUS and US surveys; however, this difference remained significant for the CAN and UK surveys as well ($P < 0.001$ in both cases). c-Statistic describing the area under the ROC curve was 0.80 (C-Spine) and 0.76 (CT-Head), indicating good discrimination between those who would and would not consider using the rules.

Hypotheses 4 and 5: Region and Rule

Figure 1 depicts mean instrument scores across the 4 different regions and for the 2 decision rules. ANOVA showed a significant effect of Rule ($F_{1,1234} = 61.3$, $P < 0.001$) resulting from generally higher acceptability scores (averaged across all 4 countries) for C-Spine over CT-Head. A significant effect of Region ($F_{3,1234} = 42.4$, $P < 0.001$) shows that scores differed between regions; further post hoc t test analyses (with Bonferroni adjusted significance levels) showed that acceptability across both rules was highest in Canada. Furthermore, a significant Rule × Region interaction ($F_{3,1234} = 28.5$, $P < 0.001$) stems from different regions showing different instrument scores for the 2 rules. As can be seen in Figure 1, physicians from the UK rated C-Spine considerably more highly than CT-Head; in the US, this difference was smaller, and there was little or no difference for Canada and Australasia.

Hypothesis 5: Construct Coverage

Physicians who indicated that they would not consider using a rule were asked to indicate why through an open-ended question. Two raters independently evaluated whether these comments were addressed by 1 or more of the 12 instrument items. For the C-Spine, 183 comments were given (all regions included). Both raters indicated that the majority of comments were addressed by 1 or more of the instrument items (86.9% and 91.3% of comments). Raters agreed on the specific item with which a comment was associated for 90.2% of the instrument-relevant comments. For the CT-Head, 228 comments were given. Again, both raters indicated the majority of comments were addressed by 1 or more items in the instrument (82.9% and 86% of comments), and raters agreed on the instrument item for 85.5% of the comments. Cohen's κ across both rules was 0.82. Comments were assigned to all 12 items of the instrument, suggesting no item should be dropped for lack of relevance. Analysis of the comments that were not relevant to the instrument, perhaps to develop new items to append to the OADRI, will be the subject of future work.

DISCUSSION

We designed the OADRI to measure acceptability of clinical decision rules and to serve as a predictor of future use of such rules. Our study findings support the validity of our instrument. First, acceptability scores were higher among clinicians who reported using the rules than among nonusers. Second, among those who do not already use the rules, acceptability scores were higher among those who would consider using a rule in the future than among than those who would not consider using it. Third, among those who do currently use the rules, acceptability scores were higher among those who report using it ''all of the time'' compared with those who report using the rule less often.

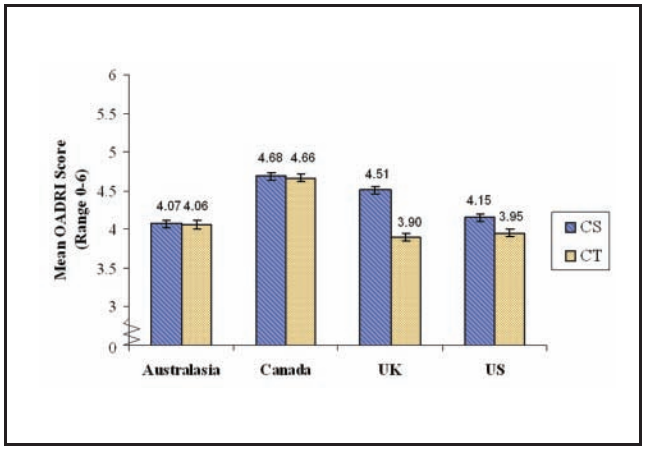

Figure 1 Mean OADRI scores (SE) across the four regions and for both rules. Based on 1238 respondents who completed the OADRI for both rules. Scores are generally higher for C-Spine than CT-Head (main effect of rule), generally higher in Canada than for the other regions (main effect of region), and the difference in acceptability between C-Spine and CT-Head is bigger in the UK than in the other regions (interaction).

Acceptability scores differentiated between those who say they use the rule ''all of the time,'' those who said they use it ''most of the time,'' and those who say they use it only ''sometimes.'' Fourth, acceptability scores overall were higher for C-Spine than CT-Head, as predicted. Finally, acceptability scores differed according to region, with ratings from the region of origin higher than in any other region. These tests of construct validity provide strong evidence that the OADRI performed as intended and discriminates between relevant subgroups of responders. Furthermore, the instrument is brief, easy-to-complete, and can be administered regardless of respondents' familiarity with the target rule.

Streiner[36] draws a distinction between 2 kinds of questionnaire: scales and indices. Scales consist of correlated items that are thought to be caused by a single, underlying construct. For example, responses to scale items measuring mathematical ability, problem-solving ability, and so forth are considered to be caused by an underlying construct of intelligence. Indexes, in contrast, consist of often uncorrelated items that are better thought of as potential causes of a construct rather than being caused by it. For example, index items measuring mobility, feeding, leisure activities, and so forth are better thought of as potential causes of the construct of quality of life rather than caused by it. Streiner further notes that correlations between items in an index may change considerably from 1 group to another, depending on which factors cause the construct within the group. As a result of this contextual independence of index items, some psychometric tools usually applied to scales, such as measures of internal consistency and factor analysis, need to be interpreted with great care and, in some cases, may not be appropriate at all.

The OADRI was initially conceived as a scale for measuring decision rule acceptability. As a scale, and for these groups, it showed high internal consistency and high construct validity across the populations surveyed. However, consistent with Streiner's characterization of the scale/index distinction as a continuum rather than a dichotomy, we note that the OADRI has characteristics of an index as well; many OADRI items (e.g., easy to use) seem better described as causing, rather than being caused by, rule acceptability. Furthermore, the items have the potential to behave independently. For example, if a clinician believed that use of the rule would increase their legal liability, it likely would not matter that the rule was easy to use or remember; the rule would be considered unacceptable on that basis alone. In fact, this effect seems to have manifested itself in some relatively low item-total correlations for the ''lawsuits'' item in Table 1.

This notion that the OADRI may serve both as a scale and as an index has several implications. First, it suggests that analysis of the underlying dimensionality of acceptability based on these items (e.g., factor analysis) may often be less fruitful than examination of outlier items, because the extent to which items ''hang together'' as 1 or more factors will vary between contexts. We conducted exploratory factor analyses for both C-Spine and CT-Head responses and found inconsistent, difficult-to-interpret factor structures. However, as discussed above, examination of individual items can yield important hypotheses about barriers to acceptability among specific subgroups. Second, in homogeneous subgroups where an individual item may behave independently of the others, information about the specific barriers to rule acceptability may be more important than the overall instrument score. In the example cited above, instrument scores would decrease only to the extent that the ''lawsuits'' item can affect the overall mean and, thus, not sensitively measure changes in acceptability. Third, because the OADRI has characteristics of an index, choice of the specific items to ensure construct coverage becomes particularly important. In the current context, the OADRI was shown to be a valid measure of acceptability. In other contexts, where barriers not included

by OADRI items might be seen as important, addition of other index items might be warranted.

Keeping this distinction in mind, we see the OADRI serving 2 important roles. First, instrument scores will allow comparison of the overall acceptability of different rules. This may involve using the OADRI to decide which of several ''proto-rules'' should be targeted for further development, or evaluating multiple existing rules to see which will be more successful when implemented in a new environment. Second, analysis of individual items will allow identification of specific barriers to rule acceptability. After identifying these specific barriers, rule developers can decide how to address these barriers, perhaps through rule modification or communication strategies specifically targeting the barrier. Information about barriers can also inform the design of knowledge translation activities for those wishing to implement a rule into a new clinical setting.

Despite these encouraging initial data, measuring a complex construct such as acceptability requires ongoing testing.[28] We would expect, for example, acceptability scores to vary between rules that differ in the methodological rigor of their development. Assuming respondents know about and understand these differences, instrument scores should be higher for a higher quality rule. The current study could not test this hypothesis, because the 2 target rules underwent similar rigorous development processes. The methods used in the current study also prevent us from evaluating whether more complex rules are less acceptable than simpler rules,[1,2] whether some physicians' predispositions against rules in general results in lower scores,[3,4] and whether specific environmental barriers to use of rules affects physician scores.[37] Further validation of the OADRI will be required to establish the ability of the instrument to identify differences in acceptability due to these factors.

LIMITATIONS

This study has limitations that warrant consideration. First, as mentioned above, the OADRI requires validation on a larger range of rules drawn from a greater range of clinical disciplines. We validated the OADRI using 2 Canadian emergency medicine rules that were developed to the same rigorous standards. Despite the similarities, the OADRI showed differences in acceptability between these rules. This suggests that the OADRI would likely be capable of discriminating between the acceptability of more disparate rules, but this must be tested empirically.

Second, response rates to our surveys range fell in the 40% to 70% range, suggesting a need to consider whether sampling bias affected our results. If nonresponders chose not to respond because of a lack of interest in clinical decision rules, a smaller proportion of the population might report using rules than was estimated in our sample. Furthermore, overall acceptability scores of our samples might be overestimated compared with the population. However, a very similar proportion of respondents for the C-Spine and CT-Head suggests that bias will likely be minimal in comparisons between the 2 rules. Furthermore, if our sample is relatively homogeneous, estimates of group differences based on the validation hypotheses (e.g., higher OADRI scores for those who use the rule consistently) would, if anything, be underestimates of the population-derived scores.

Third, we have not yet examined the performance of the OADRI over time. An evaluation of the stability of scores for individual raters over relatively short periods of time (i.e., test-retest reliability) would be useful and is a priority for future work. As well, it will be essential to evaluate the responsiveness of the OADRI to changes in acceptability over longer time periods, as may occur when successful knowledge translation strategies are implemented. An understanding of how well the OADRI detects changes in acceptability overall, and how well it detects changes in perceived barriers at different phases of the knowledge translation cycle, will greatly increase the utility of the instrument.

Fourth, constraints imposed by the national physician organizations required us to alter our survey protocols between the 4 regions. We cannot therefore be certain whether measured regional differences (e.g., in response rates) stem from the region or the survey protocol. However, we note that our findings primarily indicate that the OADRI performed consistently across all 4 regions and both target rules.

Fifth, we did not randomize or counterbalance order of presentation of the 2 iterations of the OADRI; the C-Spine was always evaluated first followed by the CT-Head. Presentation order therefore may have contributed to the slightly reduced number of responses to the CT-Head and altered the point estimates of acceptability for the 2 rules. However, the small difference in OADRI response rate to the 2 rules (C-Spine, 98.7%; CT-Head, 95.2%), and the fact that acceptability scores behaved similarly for the 2 rules across a range of hypotheses, suggests

to us that order effects were likely relatively unimportant in this context.

Finally, because of the cross-sectional nature of the current study, we may have obtained inflated correlations between instrument scores and stated rule use. Although we were careful to ask respondents about their current use of the C-Spine and CT-Head rules before administering the OADRI, the correlation may have been inflated as respondents strived for consistency between different sections of the survey. (''I've said I use this rule, so I should score it high on these questions''). Future prospective validation should include administering the OADRI to physicians before implementation of a rule and then examining the correlation between OADRI scores and later adoption of the rule into practice. Confirming predictive validity will greatly improve utility of the OADRI, moving it beyond its current function of measuring acceptability to serving as a predictive measure of ultimate rule use.

CONCLUSIONS/RECOMMENDATIONS

Clinical decision rules can produce many benefits for clinicians, patients, health care systems, and payers, but they are expensive to develop and will only realize these benefits if they are widely used. The OADRI is a simple, 12-item instrument designed to indicate whether a proposed rule will be accepted and used by the target audience. For rule developers, the instrument may help determine which of multiple proto-rules has the best odds of acceptance, as well as suggest specific barriers to address early in the development process. For those implementing a new rule into practice, the instrument could indicate its acceptability, and perhaps serve as a core set of questions designed to identify specific barriers to use, prior to resource-intensive implementation strategies. Use of the OADRI should not only involve interpretation of the summary acceptability score, but should also consider that the instrument may exhibit properties of an index, as opposed to a scale. Future work could examine whether the OADRI can be applied to or modified for nonphysician health care providers, and should include prospective validation to examine the extent to which OADRI scores predict future use of clinical decision rules.

ACKNOWLEDGMENTS

The authors thank Kelly Carroll for help in preparation of the manuscript, 2 anonymous reviewers for constructive comments on the draft, and all of the physician respondents who volunteered their time and expertise to complete this survey.

REFERENCES

1. Brehaut JC, Stiell I, Graham I, Visentin L. Clinical decision rules 'in real world': how a widely disseminated rule is used in everyday practice. Acad Emerg Med. 2005;12:948–56.

2. Brehaut JC, Stiell I, Graham I. Will a new clinical decision rule be widely used? The case of the Canadian C-Spine Rule. Acad Emerg Med. 2006;13:413–20.

3. Graham ID, Stiell IG, Laupacis A, O'Connor AM, Wells GA. Emergency physicians' attitudes toward and use of clinical decision rules for radiography. Acad Emerg Med. 1998;5:134–40.

4. Graham ID, Stiell IG, Laupacis A, et al. Awareness and use of the Ottawa Ankle and Knee Rules in 5 countries: can publication alone be enough to change practice? Ann Emerg Med. 2001;37:259–66.

5. Bandiera G, Stiell IG, Wells GA, et al. The Canadian C-Spine Rule performs better than unstructured physician judgement. Ann Emerg Med. 2003;42:395–402.

6. Jacobs LM, Schwartz R. Prospective analysis of acute cervical spine injury: a methodology to predict injury. Ann Emerg Med. 1986;15:44–9.

7. McKee TR, Tinkoff G, Rhodes M. Asymptomatic occult cervical spine fracture: case report and review of the literature. J Trauma. 1990;30:623–6.

8. Brand DA, Frazier WK, Kohlhepp WC, et al. A protocol for selecting patients with injured extremities who need x-rays. N Engl J Med. 1982;306:333–9.

9. Dunlop MG, Beattie TF, White GK, Raab GM, Doull RI. Guidelines for selective radiological assessment of inversion ankle injuries. BMJ. 1986;293:603–5.

10. Stiell IG, Greenberg GH, McKnight RD, Nair RC, McDowell I, Worthington JR. A study to develop clinical decision rules for the use of radiography in acute ankle injuries. Ann Emerg Med. 1992;21:384–90.

11. Stiell IG, McKnight RD, Greenberg GH, et al. Implementation of the Ottawa Ankle Rules. JAMA. 1994;271:827–32.

12. Stiell IG, Greenberg GH, Wells GA, et al. Derivation of a decision rule for the use of radiography in acute knee injuries. Ann Emerg Med. 1995;26:405–13.

13. Stiell IG, Wells GA, Hoag RH, et al. Implementation of the Ottawa Knee Rule for the use of radiography in acute knee injuries. JAMA. 1997;278:2075–9.

14. Stiell IG, Wells GA, Vandemheen KL, et al. The Canadian C-spine rule for radiography in alert and stable trauma patients. JAMA. 2001;286:1841–8.

15. Stiell IG, Wells GA, Vandemheen K, et al. The Canadian CT Head Rule for patients with minor head injury. Lancet. 2001;357:1391–6.

16. Stiell IG, Clement CM, McKnight RD, et al. The Canadian C-spine rule versus the NEXUS low-risk criteria in patients with trauma. N Engl J Med. 2003;349:2510–8.

17. Laupacis A, Sekar N, Stiell IG. Clinical prediction rules. A review and suggested modifications of methodological standards. JAMA. 1997;277:488–94.

18. McGinn TG, Guyatt GH, Wyer PC, Naylor CD, Stiell IG, Richardson WS. Users' guides to the medical literature. XXII: How to use articles about clinical decision rules. JAMA. 2000; 284:79–84.

19. Stiell IG, Wells GA. Methodologic standards for the development of clinical decision rules in emergency medicine. Ann Emerg Med. 1999;33:437–47.

20. Stiell I, Clement CM, Rowe BH, et al. Comparison of the Canadian CT Head Rule and the New Orleans Criteria in patients with minor head injury. JAMA. 2005;294:1511–8.

21. Stiell IG, Lesiuk H, Wells GA, et al. Canadian CT head rule study for patients with minor head injury: methodology for phase II (validation and economic analysis. Ann Emerg Med. 2001;38: 317–22.

22. Stiell IG, Clement C, Grimshaw J, Brehaut J, et al. Implementation of the Canadian C-Spine Rule: a prospective 12-centre cluster randomized trial. BMJ, in press.

23. Stiell IG, Wells G, Laupacis A, et al. Multicentre trial to introduce the Ottawa Ankle Rules for use of radiography in acute ankle injuries. BMJ. 1995;311:594–7.

24. Reilly B, Evans JS. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. Ann Intern Med. 2006;144:201–9.

25. Brehaut JC, Stiell I. Clinical sensibility and barriers to KT should be considered prior to the impact analysis. Ann Intern Med. 2006;77–8.

26. Cabana M, Rushton J, Rush A. Implementing practice guidelines for depression: applying a new framework to an old problem. Gen Hosp Psychiatry. 2002;24:35–42.

27. Hayward RSA, Guyatt GH, Moore KA, McKibbon A, Carter AO. Canadian physicians' attitudes about and preferences regarding clinical practice guidelines. CMAJ. 1997;156:1715–23.

28. Streiner D, Norman G. Health Measurement Scales: A Practical Guide to Their Development and Use. New York: Oxford University Press; 1995.

29. Eagles D, Stiell I, Clement C, et al. An international survey of emergency physicians' awareness and use of the Canadian Cervical-Spine Rule and the Canadian Computed Tomography Head Rule. Acad Emerg Med. 2008;15:1256–1261.

30. Brouwers MC, Chambers A, Perry J. Can surveying practitioners about their practices help identify priority clinical practice guideline topics? BMC Health Serv Res. 2003;3:23.

31. Graham I, Beardall S, Carter A, Tetroe J, Davies M. The state of the science and art of practice guideline development, dissemination, and evaluation in Canada. J Eval Clin Pract. 2003;9:195–202.

32. Brouwers MD, Graham ID, Hanna SE, Cameron DA, Browman GP. Clinician's assessments of practice guidelines in oncology: the CAPGO survey. Int J Technol Assess Health Care. 2004;20:421–6.

33. Graham I, Logan J. Translating research: innovations in knowledge transfer and continuity of care. Can J Nurs Res. 2004; 36:89–103.

34. Logan J, Graham ID. Toward a comprehensive interdisciplinary model of health care research use. Sci Commun. 1998;20:227–46.

35. Dillman D. Mail and internet surveys: The Tailored Design. New York: John Wiley & Sons; 2000.

36. Streiner DL. Being inconsistent about consistency: when coefficient alpha does and doesn't matter. J Pers Assess. 2003;80:217–22.

37. Dayan PS, Osmond M, Kuppermann N, et al. Development of the capacity necessary to perform and promote knowledge translation research in emergency medicine. Acad Emerg Med. 2007; 14:978–83.

38. Brehaut JC, Graham I, Stiell I, Visentin L. Print Format and sender recognition are related to survey completion rate, and maybe response rate. J Clin Epidemiol. 2006;59:635–41.

39. Kerr D, Bradshaw L, Kelly AM. Implementation of the Canadian C-spine rule reduces cervical spine x-ray rate for alert patients with potential neck injury. J Emerg Med. 2005;28:127–31.