Developing priority criteria for magnetic resonance imaging: results from the Western Canada Waiting List Project

David C. Hadorn, MD, PhD, and the Steering Committee of the Western Canada Waiting List Project*

Objective: The Western Canada Waiting List (WCWL) Project is a federally funded partnership of 19 organizations, including medical associations, health authorities, ministries of health and research organizations, that was created to develop tools to assist in assessing the relative urgency and priority of patients on waiting lists. The WCWL panel on magnetic resonance imaging (MRI) was 1 of 5 panels constituted under this project. **Methods:** The panel developed and tested a set of standardized clinical criteria for setting priorities among patients awaiting MRI. The criteria were applied to 407 patients in the 4 western provinces. Regression analysis was used to determine the set of criteria weights that collectively best predicted clinicians' overall ratings of patients' urgency for MRI. Reliability was assessed using clinicians' ratings of 6 hypothetical paper cases. **Results:** The resulting weighted criteria accounted for about two-fifths of the observed variance in overall urgency ratings (R^2 = 39.9%). The panel then modified the criteria on the basis of regression results and clinical judgment. Most of the revised criteria items showed poor inter-rater reliability, but test–retest reliability (over a 2-month interval) was relatively good. **Conclusion:** Criteria items requiring probability judgments were a challenge for clinicians. Further development and testing of the tool appears warranted, although considerable question remains concerning the utility of priority criteria for MRI and other diagnostic services.

Objectif : Le Projet sur les listes d'attente dans l'Ouest du Canada est un partenariat subventionné par le fédéral qui réunit 19 organisations dont des sociétés médicales, des administrations de santé, des ministères de la santé et des organismes de recherche. Son rôle est de trouver des façons pratiques d'évaluer le degré d'urgence et la priorité des patients en liste d'attente. Le comité de l'imagerie par résonance magnétique (IRM) est l'un des cinq comités de travail constitués en vertu de ce projet. **Méthodes** : Le comité a mis au point et vérifié un ensemble de critères cliniques standardisés pour établir des priorités parmi les patients en attente d'IRM. On a utilisé ces critères chez 407 patients répartis dans les 4 provinces de l'ouest. Par analyse de régression, on a l'ensemble de critères qui correspondait le mieux à ceux suggérés par les médecins traitants quant à la priorité de l'IRM. La fiabilité fut jugée en fonction de l'opinion des médecins sur six cas de présentations hypothétiques. **Résultats** : Les critères pondérés retenus ont compté pour les deux cinquièmes environ de la variance observée dans l'ensemble des évaluations d'urgence (R^2 = 39,9 %). Le comité a par la suite modifié les critères selon les résultats de l'analyse de régression et du jugement clinique. La plupart des critères révisés montrèrent peu de fiabilité entre eux, mais la fiabilité test-retest (sur une période de 2 mois) fut relativement bonne. **Conclusion** : La mise au point de critères faisant appel à des jugements de probabilité fut un grand défi pour les médecins. La poursuite du travail de perfectionnement et de mise à l'essai de la méthode semble justifiée, bien qu'il reste beaucoup de questions à résoudre quant à l'utilité de critères de priorité pour l'IRM et les autres services diagnostics.

Waiting lists for health care services are common in all publicly funded national health services. They are a regular source of public distress and political consternation, in large part because they entail extended suffering, disability and, occasionally, death for patients on those lists.

*The complete list of members of the Western Canada Waiting List Project Steering Committee appears at the end of the article.

Address for correspondence: Dr. Tom Noseworthy, Professor and Director, Centre for Health Research and Policy Studies, Rm. 21, Heritage Medical Research Bldg., University of Calgary, Calgary AB T2N 4N1; fax 403 210-9378; tnosewor@ucalgary.ca

Submitted Aug. 8, 2001
Revision requested Nov. 23, 2001
Resubmitted Dec. 14, 2001
Accepted Dec. 14, 2001

Can Assoc Radiol J 2002;53(4):210-8.

Equally troublesome is the prevalent impression that waiting lists may not be fair.[1,2] A growing body of evidence indicates that patients' chances of receiving needed services in a timely manner, based solely on clinical urgency, are uncertain at best. A recent report concluded that the management of waiting lists across Canada is, in general, "chaotic," as well as "non-standardized, capriciously organized, poorly monitored, and … in grave need of retooling."[3] As such, the authors conclude it is "impossible to … rationally manage the patients on those lists."[3] And impossible, therefore, to guarantee fairness.

Similar concerns underlay New Zealand doctors' participation in developing point-count priority criteria for assessing urgency for a variety of clinical conditions, despite significant initial misgivings. Clinicians "almost universally acknowledged that decisions about urgency and priority were made inconsistently. Often, the "squeaky wheel would get the grease," and more deserving but uncomplaining patients would be disadvantaged."[4]

PRIORITY CRITERIA

In response to the need for better waiting list management, an increasing number of clinicians and health authorities are adopting point-count measures for assessing patients' relative clinical urgency or priority.[5] Similar point-count measures are used in many settings throughout medicine to assess severity of illness and risk of adverse events (e.g., Apgar score to assess neonatal stability and APACHE score to estimate the probability of dying in intensive care units). Such measures function as additive or linear models from a statistical point of view.[6]

Similarly, priority criteria estimate severity of illness as an indicator of urgency, although additional considerations are often incorporated in criteria, including such social and role issues as whether patients' illnesses are interfering with their ability to care for themselves or their households. The principal functions of priority criteria are: (1) to guide decisions about the relative urgency and order of patients on waiting lists and (2) to develop case-mix descriptions of patients on waiting lists. These descriptions can be used to assess and compare waiting lists across regions and over time.

Initial experience with priority criteria in New Zealand, Sweden, England, Northern Ireland and Wales was reviewed in a recent report prepared by the Health Policy and Economic Research Unit of the British Medical Association.[7] The report endorsed the following approach:

> The widespread introduction, in this country (England), of some form of priority scoring system for surgical waiting lists would have benefits, e.g., greater transparency for all (patients, managers, politicians and doctors) on the decision taken to offer surgery to a patient; in the long term a system that is equitable across the whole country; provision of service led by clinical need and in the control of clinicians.[7]

Little information has been published concerning the clinical validity of priority criteria, however. In New Zealand, regression analysis was used to generate weights for sets of criteria on the basis of a comparison with overall clinical judgment.[8] However, the number of patients included in most of these analyses was relatively low (e.g., 97 cataract patients). Only coronary artery bypass graft surgery, with 260 patients, assessed a sufficient number of patients to permit calculation of stable criteria weights.

Similarly, little or no effort has been made to assess the intra- or interobserver reliability of priority criteria (i.e., the extent to which raters arrive at the same [or similar] ratings using the criteria when evaluating the same [or similar] patients). A study conducted in the mid-Hampshire region of England found good agreement between the priority scores assigned by general practitioners and specialists to patients with hip or knee arthritis or cataracts.[9]

WESTERN CANADA WAITING LIST PROJECT

The Western Canada Waiting List (WCWL) Project was established with a grant from Health Canada's Health Transition Fund to address some of the problems in waiting list management identified in the report by McDonand et al.[3] In particular, the project focused on developing, testing and refining clinical measures for assessing and comparing the relative urgency of patients on waiting lists.

The WCWL Project was a collaborative undertaking by 19 partner organizations: 7 regional health authorities, 4 medical associations, 4 provincial ministries of health and 4 health research centres. Clinical panels consisting of specialists, family physicians and other relevant health care providers were constituted to address each of 5 areas: cataract surgery, general surgery, hip and knee replacement, magnetic resonance imaging (MRI), and children's mental health services. This article describes the experience of the MRI panel.

Most systems currently used to categorize patients by urgency for MRI use the time-honoured method of "emergency, urgent, semi-urgent or routine" and its variants. Such systems are highly subjective and probably inadequate to describe the urgency case-mix of patients on MRI waiting lists. Whether a more sophisticated system, consisting of priority criteria and scores, can be made to work in this setting is unknown. To our knowledge, this project represents the

first effort to evaluate the relative urgency of MRI for patients on waiting lists using explicit measures (i.e., priority criteria).

Because MRI is a diagnostic procedure, the task of assessing relative urgency is complicated by the need to estimate the likelihood that MRI will provide critical diagnostic information, as well as the likelihood that subsequent treatment will improve health outcomes. This adds substantial uncertainty and complexity to the task of assessing the extent to which patients are likely to experience improved health outcomes (e.g., improved longevity, decreased pain) as a result of receiving an MRI examination.

METHODS

The MRI panel was composed of 7 academic and community radiologists, 1 family physician, a neurologist, a pediatric neurologist, an orthopedic surgeon, a neurosurgeon and 2 health administrators. Panelists were drawn from all 4 western provinces. The panel was in place from October 1999 through June 2000 and met formally 3 times during that period, with considerable work done between meetings.

At its first meeting, the panel identified the clinical, functional and social factors they considered when assessing the relative urgency of patients on waiting lists for MRI. The factors designed to reflect severity of illness were similar to criteria developed as part of the New Zealand program.[4]

On the basis of this discussion, the panel adopted a set of criteria incorporating both indicators of patients' severity of illness and estimates of the likely clinical value of the MRI in terms of improved health outcomes. The panel decided to develop priority criteria that would cover the entire case-mix of patients seeking MRI evaluation. Although this approach presented difficulties in terms of the broad range of diagnoses and age groups covered, it was acknowledged that radiologists currently set priorities across all patients, so any instrument intended to assist in this task must be sufficiently broad. Similar generic, service-wide criteria were developed by the general surgery and children's mental health WCWL panels. In contrast, the criteria developed by the hip and knee replacement and cataract surgery panels were procedure-specific.

The initial criteria, shown in the first column of Table 1, were incorporated into a questionnaire, which was then used by 15 referring specialists, including 3 panelists, to rate a series of consecutive patients in their practices. Referring specialists included orthopedic surgeons, neurosurgeons and neurologists from 4 western Canadian provinces. The MRI panelists decided it would be desirable for much of the initial pilot testing to be done by these referring specialists, because these physicians were closer to patients and had a better understanding of their clinical situations. Pilot testing data collection took place from November 1999 to May 2000.

Participating physicians assigned each patient to the appropriate level on each criterion (e.g., mild pain, moderate limits in role function) and also rated the overall urgency of each patient on a 10-cm visual analogue scale (VAS). This latter rating served as the dependent variable in regression analyses, which were used to determine the statistically optimal set of weights on each criterion to best predict (or correlate with) overall urgency. The optimal scaling method of the SPSS statistical software was used for this purpose. Standard linear regression analyses were conducted, constrained to retain all predictor variables (criteria) regardless of cross-correlations among criteria. (In standard regression, predictor variables are often dropped if they are significantly correlated with other, more highly predictive variables. However, even where significant correlations did exist among criteria, panelists wished to retain all criteria to ensure adequate face validity. Thus, it would probably be unacceptable from a clinical point of view to remove from the questionnaire explicit consideration of, for example, the extent of symptoms patients are experiencing, even if these tended to correlate with [or to be "captured by"] scores on other items.) The calculated item weights reflected the relative contribution of each criterion to the total urgency score. Univariate correlations among criteria were measured using standard Pearson correlation statistics.

The initial regression analysis was performed on the complete set of variables. A further regression analysis was conducted in which highly correlated predictors were collapsed into a smaller set of broader criteria. In this analysis, items 6 (probability of finding significant pathology) and 9 (probability of MRI providing essential diagnostic information) were combined into 1 item. Similarly, items 3, 4 and 5 (impairment in role, social and self-care functions) were amalgamated. The additional analysis was designed to investigate the impact of reducing the number of factors to make the instrument more "user friendly."

Preliminary work to evaluate the reliability of the priority criteria was undertaken in June 2000. Six paper cases were developed by one of the panel members to depict typical cases referred for MRI scanning. These cases incorporated a range of neurological and orthopedic problems and were reviewed and scored independently by 9 panel members and their colleagues. Inter-rater reliability was assessed for each of the criteria items and the VAS urgency rating using the intraclass correlation coefficient (ICC). In addition, to clarify ambiguities and suggest further refinements to the tool, a qualitative session was conducted in which panel members reviewed and rescored 2 of the

paper cases and discussed their ratings. (Only 2 cases were discussed in this level of detail because of time constraints.)

On the basis of the results of the initial pilot testing and reliability work, panellists modified the original set of criteria and adopted a set of criteria weights.

Table 1: Summary of criteria and score development for the Western Canada Waiting List Project magnetic resonance imaging (MRI) priority form

Original tool 18 Oct 1999	Revised tool 13 Mar 2000	Pilot testing analysis R^2 = 39.9%*			Tool refinement 22 Jun 2000	
Item, level	Item, level	No. (and %)		Score†	Weight	Item, level
1. Usual duration, frequency, intensity of pain	1. Usual duration, frequency, intensity of pain					1. Usual duration, frequency, intensity of pain and/or suffering
None		158	(39)	0		
Mild		83	(20)	0	0	None or mild
Moderate	No change in levels	132	(33)	1	10	Moderate
Severe		32	(8)	3	**20**	Severe
Total		405	(100)			
2. Usual duration, frequency, intensity of other forms of suffering	2. Usual duration, frequency, intensity of other forms of suffering					
None		114	(29)	0		Items 1 and 2 were combined
Mild		128	(32)	1		
Moderate	No change in levels	120	(30)	*7*		
Severe		35	(9)	11		
Total		397	(100)			
3. Degree of impairment of role function‡	3. Degree of impairment of role function‡					2. Severity of illness, impairment
None		103	(25)	0		
Mild		118	(29)	2	*5*	Low
Moderate	No change in levels	132	(33)	4	10	Medium
Severe		51	(13)	*5*	**20**	High
Total		404	(100)			
4. Degree of impairment of social activity or participation‡	4. Degree of impairment of social activity or participation‡					
None		96	(24)	0		Items 3, 4, 5, 7 were combined
Mild		121	(30)	0		
Moderate	No change in levels	148	(36)	0		
Severe		41	(10)	1		
Total		406	(100)			
5. Problems with self-care‡	5. Problems with self-care‡					
None		264	(66)	0		
Mild		78	(19)	0		Items 3, 4, 5, 7 were combined
Moderate	No change in levels	44	(11)	0		
Severe		15	(4)	1		
Total		401	(100)			
6. Probability of finding significant pathology	6. Probability of finding significant pathology					3. Probability of MRI providing clinically significant diagnostic information
Low		87	(22)	0	*5*	Low
Medium		137	(34)	1	10	Medium
High	No change in levels	111	(27)	1	**20**	High
Very high		70	(17)	1		
Total		405	(100)			

(continued)

Table 1: (cont. from previous page) Summary of criteria and score development for the Western Canada Waiting List Project magnetic resonance imaging (MRI) priority form

Original tool 18 Oct 1999	Revised tool 13 Mar 2000	Pilot testing analysis R^2 = 39.9%*			Tool refinement 22 Jun 2000	
Item, level	Item, level	No. (and %)		Score†	Weight	Item, level
7. Probable level of morbidity	7. Probable level of morbidity associated with suspected pathology					
Low		128	(32)	0		Items 3, 4, 5, 7 were combined
Medium		170	(42)	4		
High	No change in levels	77	(19)	6		
Very high		29	(7)	10		
Total		404	(100)			
8. Probable time course for morbidity	8. Probable time course for clinical deterioration					4. Probable time course of clinical deterioration
	Not applicable					
Long-term	Long-term	191	(51)	0	5	Long-term (> 6 mo)
Near-term	Short-term	131	(35)	29	10	Mid-term (1–6 mo)
Immediate	Immediate	54	(14)	34	**20**	Short-term (< 1 mo)
Total		376	(100)			
9. Probability of MRI providing essential treatment-altering information	9. Probability of MRI providing essential diagnostic information					
Low		87	(22)	0		Items 6 and 9 were combined
Medium		151	(37)	5		
High	No change in levels	119	(29)	9		
Very high		48	(12)	10		
Total		405	(100)			
10. Probability of successful intervention resulting from this information	10. Probability of successful intervention resulting from this information					5. Probability of successful treatment resulting from the diagnostic information
Low		124	(33)	0	5	Low
Medium		135	(36)	10	10	Medium
High	No change in levels	92	(24)	16	**20**	High
Very high		26	(7)	24		
Total		377	(100)			

*R^2 = 36.5% with items 3, 4 and 5 combined, and items 6 and 9 combined.
†Modelled scores inferred for groupings of related variables: patient condition (items 1, 2) = 14; patient condition (items 3–5, 7) = 17; diagnostic value (items 6, 9) = 11; natural history with no intervention (item 8) = 34; likely efficacy of treatment (item 10) = 24.
‡Items 3, 4 and 5 included additional descriptors for patients aged 5 years or younger. These were dropped in the June 2000 refinements.

Weights were apportioned among items in accordance with their relative importance, with these values scaled so that the total maximum achievable score, summing across the most severe response category for each criterion, was 100 points.

Further empirical work was conducted to evaluate both the inter-rater and test–retest (intra-rater) reliability of the revised priority criteria. Neurologists and neurosurgeons from 3 western provinces were engaged as raters because the paper cases had a predominately neuroscience focus and a high proportion of MRI referrals originate from these specialties. Otherwise, the methods were similar to those described for the preliminary reliability work. The first wave of reliability data was collected in November 2000, and retest data were obtained from 7 of these participants in January 2001. The paper cases were identical for each wave, but the presentation order was changed to control for order effects.

RESULTS

Table 1 provides a summary of criteria and score development. An interim analysis of 217 patients was prepared and discussed at the panel's second meeting in March 2000. On the basis of this analysis and panelists' experience to date (including feedback from participating referral clinicians), a series of mostly

minor changes were made in criteria wording. Perhaps the most significant change occurred in item 9, which originally asked clinicians to estimate the probability that information obtained from MRI would alter treatment plans from what would have be pursued in the absence of MRI. Although this question might reasonably be considered the "gold standard" for determining the need for MRI, at its second meeting the panel decided to drop the "treatment-altering" wording; experience suggested that clinicians often lacked sufficient information to make an informed judgment regarding this probability. Accordingly, the wording of item 9 was changed to the probability that MRI would provide "essential diagnostic information."

By the end of the data collection period, a total of 411 criteria forms were submitted, of which 407 contained useable data. Item weights were calculated, depicting the relative emphasis given to items by referring specialists when rating urgency of patients for MRI. Optimal weights (scores) inferred from regression analysis are indicated in Table 1; the R^2 value of 39.9% means that the set of criteria accounted for two-fifths of the variance in clinicians' overall urgency ratings. The highest scores were assigned to "probable time course for clinical deterioration" (revised item 8, 34%) and "probability that successful intervention would result from MRI information" (revised item 10, 24%).

Univariate correlations among criteria are presented in Table 2. A relatively high correlation (0.67) occurred between items 6 and 9, "probability of finding significant pathology" and "probability of MRI providing essential diagnostic information," respectively. Items 3, 4 and 5 had intercorrelations between 0.48 and 0.59, and item 7 had moderate correlation with these variables.

At its final meeting in June 2000, the panel made several additional changes in criteria; these are indicated in the final column of Table 1. To make the questionnaire easier to complete, the panel reduced the number of criteria from 10 to 5 on the basis of statistical correlations, conceptual similarity and evidence from preliminary reliability work. For example, item 6 was merged with item 9 because of the high statistical correlation reported above and the overlapping conceptual domains of the 2 criteria.

Weighting of the resulting 5 criteria was guided by the empirical regression results and considerations of clinical face validity. In particular, panellists felt that the clinical face validity of the scoring would be improved by increasing the maximum score available for the diagnostic value of MRI (i.e., combined items 6 and 9) above what was inferred using regression analysis (see footnote of Table 1). The final column of Table 1 summarizes the panel's revised weights. Forty percent of the maximum score was allocated to current symptom status (revised items 1 and 2); the remaining 60% was allocated evenly among diagnostic value, natural history (or risk of clinical deterioration) and likely efficacy of treatment.

A consistent theme that emerged from the pilot study was that participating clinicians found it very difficult to estimate many of the probabilities required by the criteria. It was felt that this concern was serious enough to jeopardize the overall validity and utility of the criteria set.

Further empirical work was undertaken to assess the reliability of the revised priority criteria instrument (see Table 3). There was poor inter-rater agreement among the 7 raters (neurologist and neurosurgeons) who reviewed the paper cases in January 2001 for the VAS urgency ratings for the 6 patients (ICC = 0.38). There was fair to good agreement for item 1 (ICC = 0.60) and poor agreement (ICC < 0.40) for the remaining 4 items. These ICC values should be interpreted with caution, however, because of the relatively small number of raters and minimal amount of

Table 2: Correlation matrix for MRI priority criteria using pilot testing data from 407 cases

		Item no.								
Item no.*	VAS	1	2	3	4	5	6	7	8	9
1	0.22									
2	0.25	0.16								
3	0.26	0.43	0.38							
4	0.21	0.37	0.43	0.55						
5	0.22	0.24	0.52	0.48	0.59					
6	0.33	0.19	0.19	0.29	0.19	0.27				
7	0.40	0.16	0.35	0.41	0.37	0.45	0.46			
8	0.43	0.18	0.24	0.22	0.09	0.14	0.10	0.24		
9	0.47	0.21	0.13	0.21	0.22	0.23	0.67	0.45	0.17	
10	0.52	0.33	0.17	0.31	0.29	0.30	0.56	0.39	0.18	0.71

Note: VAS = Visual Analogue Scale (dependent variable).
*See Table 1 for descriptions of items 1 through 10.

variance for some raters. Test–retest (intra-rater) reliability, measured over a 2-month interval, appeared stronger, with 1 item in the excellent range (ICC > 0.75) and the remaining items having fair to good intra-rater agreement (ICC = 0.50–0.75).

DISCUSSION

Developing clinically acceptable priority criteria for MRI is substantially more complicated than developing analogous criteria for surgical procedures or children's mental health. The primary obstacle lies in the nature of diagnostic procedures. (Of course, these difficulties were foreseen by WCWL organizers, and, indeed, MRI was selected precisely to test whether standardized clinical criteria of this sort could be developed in this context.) Whereas most wait-listed surgical procedures are applied to patients with established diagnoses and with reasonably good prospects for obtaining substantial benefit, MRI is primarily used to establish (or to rule out) conditions and diseases; moreover, many of the diagnoses commonly made using MRI have no satisfactory treatment, notably malignant brain tumours and multiple sclerosis. Regarding this latter point, most panellists felt that the value of making such diagnoses still warranted provision of MRI, although the degree of urgency that should be assigned to these cases was less clearly agreed upon. It is these extra layers of diagnostic and therapeutic uncertainty that make the task of estimating relative urgency more complicated and difficult than in surgical contexts.

There are fewer probability-based criteria in other WCWL priority instruments. The general surgery panel developed 2 life expectancy criteria (i.e., with vs. without surgery), and the hip and knee replacement panel developed a "probability of progression" criterion. Two of the 17 criteria for children's mental health involved probability assessments. By contrast, fully half of the original MRI criteria (i.e., 5 of 10) required probability judgments. Of these, 3 criteria (items 6–8) concerned the likely severity of the patient's illness (including prospect of deterioration), which is often unknown before MRI. The remaining 2 criteria (items 9 and 10) concerned the probability that MRI would provide treatment-altering (later essential) information and that subsequent treatment would be successful. In the revised set, 3 of 5 criteria ask for probability estimates.

It could be argued that radiologists who manage MRI waiting lists are already making probability estimates of the kind described above, albeit in a subjective and imprecise way, when determining patients' relative priority for MRI. Nevertheless, the requirement for explicitness entailed by the MRI probability criteria poses an extremely difficult challenge for clinicians, in view of the inevitable clinical complexities and uncertainties. On the basis of experience in the pilot test, clinicians may resist using the WCWL MRI criteria. However, removing probability estimates from the MRI criteria would severely limit their ability to reflect the likely benefit of MRI.

Significant revisions were made to the criteria at the panel's final meeting. From the panellists' perspective, the resulting criteria and assigned weights passed the test of clinical face validity. The explanatory power of the criteria in predicting judgments of overall urgency was less than might be hoped for — and less than most of the other WCWL criteria sets — but the observed degree of explanatory power was statistically significant, comparing favourably to analogous models in other medical contexts.

The degree of reliability observed across raters was, again, less than might be desired and among the lowest observed in the WCWL criteria. This is another reflection of the difficulties in making judgments associated with probability estimates and the multi-level decisions required to score the probability criteria. The stronger test–retest reliability suggests that individual raters maintain relatively consistent scoring frame-

Table 3: Inter-rater and test–retest reliability among 7 neurospecialists who rated the revised priority criteria for MRI

Item	Inter-rater reliability (Jan., 2001), ICC	Test–retest reliability,* ICC
1. Usual duration, frequency, intensity of pain and/or suffering	0.60	0.75
2. Severity of illness or impairment	0.34	0.69
3. Probable time course of clinical deterioration	0.11	0.50
4. Probability of providing clinically significant diagnostic information	0.00†	0.76
5. Probability of successful treatment from diagnostic information	0.00‡	0.62
6. Visual Analogue Scale urgency	0.38	0.68

Note: ICC = intraclass correlation coefficient. Higher score indicates greater agreement.
*Test–retest interval was 6 weeks to 2 months.
†Interpret with caution, based on 6 raters, and 2 had no variance.
‡Interpret with caution, based on 5 raters.

works over time; further qualitative work such as focus group sessions might assist in elucidating reasons for variability in probability judgments across raters.

Assuming that the challenges described above can be overcome (especially the task of estimating probabilities), a number of operational challenges will remain with the use of priority criteria for scheduling MRI. For example, patients with relatively minor (but still significant) problems will always score lower than patients with more symptomatic serious ones. As new, high-scoring patients are seen, low-scoring patients will never reach the top of the list. This problem could be addressed by adding points to the scores of patients simply for time spent waiting. This policy would, however, lead to a different problem — patients with less severe conditions would regularly "bump" patients with more severe conditions. It was for this reason that all WCWL panels decided against incorporating time for waiting among the criteria.

Another concern, raised regularly during the project, concerned the possibility that patients and clinicians would "game the system" by virtue of knowing how the point system works. This concern is especially relevant in the MRI setting because, although radiologists are (and must be) in charge of setting priorities for MRI across all clinical specialties and diagnoses, it is the referring specialists who complete the priority forms. Might neurologists, for example, be tempted to give their patients an "extra edge" over, say, orthopedic patients, or vice versa? Such concerns need to be addressed through careful monitoring, use of standard raters or other techniques. However, it is important to remember that the current chaotic and unregulated system can in most areas be easily gamed and, yet, cannot be audited.

Despite the shortcomings, we believe that the MRI criteria developed during the WCWL project provide a reasonably solid basis upon which to build. The Calgary Regional Health Authority is currently incorporating the revised criteria into their MRI information systems in a second phase of pilot testing. Experience from this and similar efforts will, it is hoped, lead to workable criteria for promoting fairness and efficiency in the management of waiting lists for MRI. A major component of the process required to develop workable criteria will be arriving at a better method for estimating the prognostic value of MRI in terms of ultimate clinical outcomes. This will not be easy, and any such method can be developed, if at all, only through the sustained efforts of radiologists and referring specialists.

ACKNOWLEDGEMENTS: We are indebted to the 19 partner organizations from the 4 Western Canadian provinces for their ongoing support throughout the WCWL Project: British Columbia Medical Association; Capital Health Region (Victoria); Vancouver/Richmond Health Board; British Columbia Ministry of Health; University of British Columbia, Centre for Health Services and Policy Research; Alberta Medical Association; Capital Health Authority (Edmonton); Calgary Regional Health Authority; Alberta Health and Wellness; University of Alberta, Department of Public Health Sciences; Saskatchewan Medical Association; Regina Health District; Saskatoon District Health; Saskatchewan Health; Health Services Utilization and Research Commission; Winnipeg Regional Health Authority; Manitoba Health; Manitoba Centre for Health Policy and Evaluation; Canadian Medical Association.

We would like to acknowledge the members of the MRI panel who contributed to the development of the MRI priority criteria tool: Dr. George Andrew, Dr. David Hadorn, Dr. Husam Darwish, Dr. Robert Diamond, Ms. Wilma Dodge, Dr. Bevan Frizzell, Dr. Bill Hall, Dr. Robert Hollinshead, Dr. D'Arcy Lawrence, Dr. Blake McClarty, Dr. Terry Myles, Dr. Neelan Pillay, Dr. Paul Schulte, Dr. Robert Sevick and Ms. Barbara Young. We would also like to thank colleagues of the panel members who participated in the pilot testing and reliability work in Winnipeg, Regina, Calgary, Edmonton and Victoria.

We would like to thank Dr. N.G. Prasad, Dr. Barbara Conner-Spady, Ms. Elaine Dunn, Ms. Helen Roman-Smith, Ms. Anne-Marie Pedersen and Ms. Katerina Carastathis for their contributions to data collection and analysis and document preparation.

FUNDING: The Western Canada Waiting List Project was supported by a financial contribution from the Health Transition Fund (Health Canada) as Project NA489. The views expressed herein do not necessarily represent the official policy of federal, provincial or territorial governments.

REFERENCES

1. Lewis S, Barer ML, Sanmartin C, Sheps S, Shortt SED, McDonald PW. Ending waiting-list mismanagement: principles and practice. *CMAJ* 2000;162:1297-300. Available: www.cmaj.ca/cgi/content/full/162/9/1297

2. Sanmartin C, Shortt SE, Barer ML, Sheps S, Lewis S, McDonald PW. Waiting for medical services in Canada: lots of heat, but little light. *CMAJ* 2000;162:1305-10. Available: www.cmaj.ca/cgi/content/full/162/9/1305

3. McDonald P, Shortt S, Sanmartin C, Barer M, Lewis S, Sheps S. *Waiting lists and waiting times for health care in Canada. More management!! More money??* Ottawa: Health Canada; 1998. Summary available: www.hc-sc.gc.ca/english/media/releases/waiting_list.html

4. Hadorn DC, Holmes AC. The New Zealand Priority Criteria Project: I: Overview. *BMJ* 1997;314:131-4.

5. Hadorn DC and the Steering Committee of the Western Canada Waiting List Project. Setting priorities for waiting lists: defining our terms. *CMAJ* 2000;163(7):857-60. Available: www.cmaj.ca/cgi/content/full/163/7/857

6. Hadorn DC and the Steering Committee of the Western Canada Waiting List Project (WCWL). Setting priorities on waiting lists: point-count systems as linear models. *J Health Serv Res Policy*. In press.

7. Health Policy and Economic Research Unit. *Waiting list prioritisation scoring systems*. London: British

Medical Association; 1998. Available free of charge from BMA House, Tavistock Square, London WC1H 9JR. See also www.bma.org.uk/ap.nsf/Content/Waiting+list+prioritisation+scoring+systems

8. Hadorn DC, Holmes A. The New Zealand Priority Criteria Project. Criteria pilot tests. [Internet supplement to Hadorn DC, Holmes AC. The New Zealand Priority Criteria Project: I. Overview. *BMJ* 1997;314:131-4.] Available: http://bmj.com/cgi/content/full/314/7074/131/DC1

9. Dent TH, Cunningham SL, Hook WE, Morris RJ, Sylvester NC. The MODEL project: a scoring system to manage demand for cataract and joint replacement surgery. *Br J Gen Pract* 2001;51:917-9.

***MEMBERS OF THE WESTERN CANADA WAITING LIST PROJECT STEERING COMMITTEE**: Dr. Tom Noseworthy (Chair), Professor and Chair, Department of Public Health Sciences, University of Alberta, Edmonton, Alta.; Dr. Morris L. Barer, Director, Centre for Heal Services and Policy Research, and Professor, Department of Health Care and Epidemiology, University of British Columbia, Vancouver, BC; Dr. Charlyn Black, Co-Director, Manitoba Centre for Health Policy and Evaluation, and Associate Head and Associate Professor, Department of Community Health Sciences, University of Manitoba, Winnipeg, Man.; Ms. Lauren Donnelly, Executive Director, Acute and Emergency Services Branch, Saskatchewan Health, Regina, Sask.; Dr. David Hadorn, Research Director, Western Canada Waiting List Project; Dr. Isra Levy, Director, Health Programs, Canadian Medical Association, Ottawa, Ont.; Mr. Steven Lewis, Partner, Access Consulting, Saskatoon, Sask.; Dr. Sam Sheps, Head, Department of Health Care and Epidemiology, University of British Columbia, Vancouver, BC; Dr. Mark C. Taylor, Assistant Professor, Department of Surgery, University of Manitoba and Deputy Head, Department of Surgery, St. Boniface General Hospital, Winnipeg, Man.; Mr. Laurence Thompson, Chief Executive Officer, Health Services Utilization and Research Commission, Saskatoon, Sask.; Mr. Darrell Thomson, Director, Economics and Policy Analysis, British Columbia Medical Association, Vancouver, BC; Ms. Barbara Young, Regional Utilization Consultant, Clinical Evaluation Services, Calgary Regional Health Authority, Calgary, Alta.; Mr. John McGurran, Director, Western Canada Waiting List Project, and Lecturer, Department of Public Health Sciences, University of Toronto, Toronto, Ont.