

Development and preliminary validation of a Function Index for Trauma (FIX-IT)

Mohit Bhandari, MD, PhD*[†]
 Scott M. Wasserman, MD[‡]
 Nicole Yurgin, PhD[‡]
 Brad Petrisor, MSc, MD*
 Sheila Sprague, MSc[†]
 Ricardo E. Dent, MD[‡]

From the *Division of Orthopaedic Surgery, Department of Surgery, and the †Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ont., and ‡Amgen Inc, Thousand Oaks, Calif.

Accepted for publication
 Oct. 9, 2012

Correspondence to:
 M. Bhandari
 293 Wellington St. N, Suite 110
 Hamilton ON L8L 8E7
 bhandam@mcmaster.ca

DOI: 10.1503/cjs.004312

Background: Assessing fracture healing in clinical trials is subjective. The new Function Index for Trauma (FIX-IT) score provides a simple, standardized approach to assess weight-bearing and pain in patients with lower extremity fractures. We conducted an initial validation of the FIX-IT score.

Methods: We conducted a cross-sectional study involving 50 patients with lower extremity fractures across different stages of healing to evaluate the reliability and preliminary validity of the FIX-IT score. Patients were independently examined by 2 orthopedic surgeons, 1 orthopedic fellow, 2 orthopedic residents and 2 research coordinators. Patients also completed the Short Form-36 version 2 (SF-36v2) questionnaire, and convergent validity was tested with the SF-36v2.

Results: For interrater reliability, the intraclass correlation coefficients ranged from 0.637 to 0.915. The overall interrater reliability for the total FIX-IT score was 0.879 (95% confidence interval 0.828–0.921). The correlations between the FIX-IT score and the SF-36 ranged from 0.682 to 0.770 for the physical component summary score, from 0.681 to 0.758 for the physical function subscale, and from 0.677 to 0.786 for the role-physical subscale.

Conclusion: The FIX-IT score had high interrater agreement across multiple examiners. Moreover, FIX-IT scores correlate with the physical scores of the SF-36. Although additional research is needed to fully validate FIX-IT, our results suggest the potential for FIX-IT to be a reliable adjunctive clinician measure to evaluate healing in lower extremity fractures.

Level of evidence: Diagnostic Study Level I.

Contexte : Évaluer la guérison d'une fracture dans le cadre d'essais cliniques est un processus subjectif. Le nouveau score FIX-IT (pour Function IndeX for Trauma) constitue une approche simple et standardisée pour évaluer la mise en charge et la douleur chez les patients ayant subi une fracture d'un membre inférieur. Nous avons procédé à une validation initiale du score FIX-IT.

Méthodes : Nous avons réalisé une étude transversale regroupant 50 patients qui ont subi une fracture d'un membre inférieur, à différents stades de la guérison, pour évaluer la fiabilité et la validité préliminaire du score FIX-IT. Les patients ont été examinés indépendamment par 2 chirurgiens orthopédistes, 1 chargé de cours en orthopédie, 2 médecins résidents en orthopédie et 2 coordonnateurs de recherche. Les patients ont aussi répondu au questionnaire SF-36v2 (Short Form-36 version 2) et la validité convergente a été vérifiée au moyen du SF-36v2.

Résultats : En ce qui concerne la fiabilité interexamineur, les coefficients de corrélation intraclass ont varié de 0,637 à 0,915. La fiabilité interexamineur pour le score FIX-IT total a été de 0,879 (intervalle de confiance de 95 % 0,828–0,921). Les corrélations entre le score FIX-IT et le SF-36 ont varié de 0,682 à 0,770 pour le score sommaire de la composante physique, de 0,681 à 0,758 pour la sous-échelle du fonctionnement physique et de 0,677 à 0,786 pour la sous-échelle du rôle physique.

Conclusion : Le score FIX-IT a offert une concordance interexamineur élevée entre les multiples examinateurs. De plus, les scores FIX-IT sont en corrélation avec les scores physiques obtenus au SF-36. Même s'il faudra approfondir la recherche pour valider complètement le score FIX-IT, nos résultats donnent à penser que cet indice pourrait être une mesure clinique d'appoint fiable pour évaluer la guérison des fractures de membres inférieurs.

Niveau de preuve : Étude diagnostique de niveau I.

The clinical assessment of fracture healing is a largely subjective process without a gold standard.^{1,2} Although measures have been developed for hip and ankle fractures,³ there is no validated measure that adequately describes functional healing for tibial fractures.¹ A recently published systematic review that evaluated variability in the assessment of fracture healing in orthopedic trauma studies reported that the 3 most commonly used clinical criteria were the absence of pain or tenderness on palpation or examination, the absence of pain or tenderness when bearing weight and the ability to bear weight.⁴ Similarly, another review evaluating the clinical criteria used to define fracture union found that the 4 most common criteria were the absence of pain or tenderness when bearing weight, the absence of pain or tenderness on palpation or examination, the ability to bear weight and the ability to walk or perform activities of daily living with no pain.²

Pain at the fracture site is commonly regarded as a sign that a fracture has not yet healed.^{1,2} However, some patients may have persistent pain despite evidence of healing, whereas others may have no pain without evidence of healing.² Consequently, pain alone may be an inadequate measure to determine if a fracture has healed. The ability to bear weight on injured appendages has been suggested to serve as an objective measure for healing of tibial fractures treated by external fixation⁵ because weight-bearing ability has been shown previously to increase with time postfracture^{6,7} and has been found to correlate well with bone stiffness.⁸ In tibial shaft fractures treated with intramedullary nailing, weight-bearing is possible from the day after surgery. Therefore, early weight-bearing in this context may not represent a healed fracture, and other dimensions, such as pain, should be considered when assessing fracture healing.

The Function Index for Trauma (FIX-IT) assessment provides a simple standardized approach to the measurement of weight-bearing and pain assessment in patients with lower extremity fractures, specifically tibia and femoral fractures. The FIX-IT score is a clinical outcomes assessment measure ranging from 0 to 12 points in 2 domains: the ability to bear weight (maximum 6 points) and pain at the fracture site (maximum 6 points; see the Appendix, available at cma.ca/cjs). The ability to bear weight is assessed through the single-leg stand and ambulation procedures. Pain is assessed through palpation and stress procedures. The scores in both domains, which are weighted equally, are summed to obtain the final total score; the maximum score of 12 indicates the highest level of function. The measure was developed based on a review of published literature on the assessment of tibial fracture healing and discussion with regulatory professionals and content experts in orthopedic trauma surgery. The objective of the present study was to evaluate the face validity, content validity, external validity, overall physician satisfac-

tion, interrater reliability and convergent validity of the FIX-IT measure.

METHODS

Overview of the study design

To assess face and content validity, the FIX-IT measure was independently evaluated by 5 orthopedic trauma surgeons before the clinical study. We conducted a cross-sectional study of patients with lower extremity fractures across different stages of healing to evaluate the interrater reliability of the FIX-IT measure. We obtained research ethics board approval before initiating the study. Patients were enrolled from Hamilton Health Sciences — General Site in Hamilton, Ont., and the sample was a nonrandom, convenience sample of patients with tibial or femoral fractures presenting to a fracture clinic. To assess interrater reliability, patients were independently examined by 7 reviewers. Prior to performing the FIX-IT assessments, we recorded the demographic and fracture characteristics of patients, and the patients completed the Short-form 36, version 2 (SF-36v2) questionnaire.

Assessment of face and content validity

To assess face validity, 5 orthopedic trauma surgeons from North America, Europe and Asia independently reviewed the FIX-IT measure and determined whether it looked like it was going to measure what it was supposed to measure. Specifically, the surgeons were asked to rate on a scale of 1–5 the overall agreement with the validity of this measure for understanding functional healing in patients with fractures.

We assessed content validity qualitatively by asking each of the 5 surgeons to determine whether each item was “essential,” “useful, but not essential” or “not necessary” to the performance of the construct. Each surgeon also rated their overall satisfaction with the administration of the measure and completed an open-ended question asking whether any item essential for the performance of the construct was currently missing from the existing measure.

Assessment of interrater reliability

We assessed consecutive patients with lower extremity fractures attending a fracture clinic for inclusion in our study. The inclusion criteria were lower extremity long-bone fracture, age 18 years or older, English language, ability to ambulate before the fracture and provision of informed consent. The exclusion criteria were bilateral lower extremity fractures, fractures of the axial skeleton limiting weight-bearing, inability to complete questionnaires or comply with functional tests, presence of pre-injury lower extremity pain syndrome, paralysis or sensory

deficit and prefracture use of assistive devices. We obtained informed consent from all participants, and baseline and fracture characteristics were recorded.

Two orthopedic surgeons, 1 orthopedic fellow, 2 orthopedic surgical trainees and 2 research coordinators independently assessed patient function using the FIX-IT measure. Raters were selected before the study and each rater participated in a training session on how to use the FIX-IT measure. The team then evaluated each patient, and each rater, unaware of the other raters' responses, scored patient function in all participants.

Assessment of convergent validity

Each eligible patient completed the SF-36v2, which is a health-related quality of life measure.⁹ The SF-36v2 dimensions were scored separately and transformed to a 0–100 scale. Domains were also grouped into the physical component summary (PCS) score and the mental component summary (MCS) score, as recommended by the SF-36v2 scoring manual.⁹ We chose to use the SF-36v2 rather than other available instruments because of its use in previous studies evaluating fracture outcomes and the hypothesis that the FIX-IT measure would correlate with the SF-36v2 physical functioning scale, role-physical scale and physical health component summary measure scores.^{2,10–13}

Sample size considerations

The sample size is controlled in reliability studies by varying the number of raters and the number of patients. Although increasing the number in either group will yield a more precise reliability estimate, the number of participants has a much greater impact on the precision than the number of raters. The number of raters was chosen based on generalizability and feasibility. Using 2 orthopedic surgeons, 1 orthopedic fellow, 2 orthopedic residents and 2 research coordinators as raters, we determined that a sample of 50 patients would provide sufficient precision for meaningful analysis of the FIX-IT measure. Assuming an expected intraclass correlation (ICC) of 0.8, a sample size of 50 patients and 7 raters, the expected half width of the 95% confidence interval (CI) for the estimated ICC was approximately 0.10.¹⁴

Statistical analysis

Reviewer assessment of face and content validity and overall satisfaction with the administration of FIX-IT was assessed with 5-point Likert-type scales ranging from 1 (completely disagree) to 5 (completely agree). Scores were summarized qualitatively for each assessment.

We used ICCs with 95% CIs to measure agreement in the rater's overall FIX-IT scores, including the 4 com-

ponent scores. The ICC, used to quantify agreement for a continuous variable, is equivalent to the quadratically weighted κ for categorical data. The weighted κ , as described by Fleiss,¹⁵ adjusts the observed proportion of agreement by correction for the proportion of agreement that could have occurred by chance alone. As they are numerically equivalent, similar guidelines for interpretation of κ values can be applied to the ICC. Landis and Koch¹⁶ suggest that κ of 0–0.2 represents slight agreement, 0.21–0.40 fair agreement, 0.41–0.60 moderate agreement, and 0.61–0.80 substantial agreement. A κ value above 0.80 is considered almost perfect agreement. The value of the ICC ranges from +1, representing perfect agreement, to –1, representing absolute disagreement.

In addition, FIX-IT was compared with similar domains of a frequently used patient-reported outcomes scale. Specifically, the association between the FIX-IT

Table 1. Content and face validity of the FIX-IT measure

Question; responses	No. (%), n = 5
How would you rate your overall agreement with the validity of this measure for understanding functional healing for fracture patients?	
5 (completely satisfied)	0
4	4 (80)
3	1 (20)
2	0
1 (completely unsatisfied)	0
How satisfied are you with the administration of the measure?	
5 (completely satisfied)	0
4	3 (60)
3	2 (40)
2	0
1 (completely unsatisfied)	0
How important is the ability measured by the single-leg stand score for understanding functional healing for fracture patients?	
Essential	1 (20)
Useful, but not essential	4 (80)
Not necessary	0
How important is the ability measured by the ambulation score for understanding functional healing for fracture patients?	
Essential	3 (60)
Useful, but not essential	2 (40)
Not necessary	0
How important is tolerance measured by the palpation score for understanding functional healing for fracture patients?	
Essential	3 (60)
Useful, but not essential	2 (40)
Not necessary	0
How important is the tolerance measured by the stress score for understanding functional healing for fracture patients?	
Essential	1 (20)
Useful, but not essential	4 (80)
Not necessary	0

FIX-IT = Function IndeX for Trauma.

scores and the SF-36v2 physical functioning scale, role-physical scale, and physical health component summary measure scores were assessed using Pearson correlation. The SF-36v2 was scored according to the SF-36v2 version 2.0 scoring manuals.⁹

RESULTS

Assessment of face and content validity

The FIX-IT instrument demonstrated acceptable face and content validity, as measured by 5 experts who determined that the items were all either useful or essential (Table 1). When asked if there were any additional items to consider including in the FIX-IT measure, 2 reviewers suggested including return to work, a patient-important outcome

that indicates the ability to resume both physical and mental activities.¹⁷ As a substantial proportion of patients will not return to work even 2 years after the fracture occurs,¹⁸ return to work as a measure of fracture healing may not be very responsive to change.¹⁷ Therefore, the developers opted not to incorporate questions on return to work.

Patient characteristics

Of the 50 patients enrolled in the study, 42 (84%) had tibia fractures and 8 (16%) had femur fractures (Table 2). The mean time from injury to assessment for the study was 34 (range 0.5–555) months. The majority of the patients evaluated had already established problems with their fracture healing (Table 2).

Table 2. Participant and fracture characteristics			
Participant characteristic	No. (%) n = 50	Fracture characteristic	No. (%) n = 50
Age, yr (mean ± SD)	48 ± 16	Fracture type	
Sex		Closed	32 (64)
Male	36 (72)	Open	18 (36)
Female	14 (28)	Type of treatment	
Ethnicity		Plate fixation	22 (44)
White	40 (80)	Reamed intramedullary nail	16 (32)
Native Canadian	5 (10)	Conservative treatment	9 (18)
Hispanic	4 (8)	Screws	2 (4)
Southeast Asian	1 (2)	Nonreamed intramedullary nail	1 (2)
Time from injury, mo., mean ± SD (range)	34 ± 77 (0.5–555)	Complications	
< 6	9 (18)	None	17 (34)
6–12	4 (8)	Yes	33 (66)
12–24	19 (38)	Type of complication, n = 42‡	
> 24	18 (36)	Nonunion	13 (31)
Mechanism of injury		Delayed union	13 (31)
Fall	19 (38)	Infection	7 (17)
Motor vehicle accident	9 (18)	Wound healing problem	5 (12)
Motorcycle motor vehicle accident	6 (12)	Hardware failure	2 (5)
Pedestrian motor vehicle accident	4 (8)	Varus collapse	1 (2)
Recreational vehicle accident	4 (8)	Posttraumatic arthritis	1 (2)
Crush	3 (6)	Reoperations	
Direct trauma — blunt	3 (6)	None	22 (44)
Twist	2 (4)	Yes	28 (56)
Additional injuries		Type of reoperation, n = 47‡	
None	27 (54)	Hardware removal	13 (28)
Yes	23 (46)	Bone graft	11 (23)
Total of 33 additional injuries†		Revision open reduction internal fixation	8 (17)
Upper extremity injury	11 (33)	Irrigation and debridement	6 (13)
Other lower extremity injury	10 (30)	Sequestrectomy	4 (9)
Spinal injury	5 (15)	Skin graft	2 (4)
Head/facial injury	5 (15)	Arthrodesis	2 (4)
Thoracic injury	2 (6)	Osteotomy	1 (2)
Location of fracture		None	22 (44)
Tibia	42 (84)		
Femur	8 (16)		

SD = standard deviation.
 *Unless otherwise indicated.
 †Patients may have had more than 1 additional injury.
 ‡Patients may have had multiple complications and reoperations.

Functional status

The mean SF-36v2 PCS score was 35.06 ± 9.77 , and the mean SF-36v2 MCS score was 43.73 ± 15.55 (Fig. 1). The mean SF-36v2 physical functioning scale score was 39.32 ± 29.23 , the mean role-physical scale score was 33.92 ± 34.09 , and the mean bodily pain scale score was 42.94 ± 24.00 .

The FIX-IT measure and assessment of interrater reliability

The mean overall FIX-IT score was 7.97 ± 2.73 (Table 3). The overall interrater reliability for the FIX-IT score was

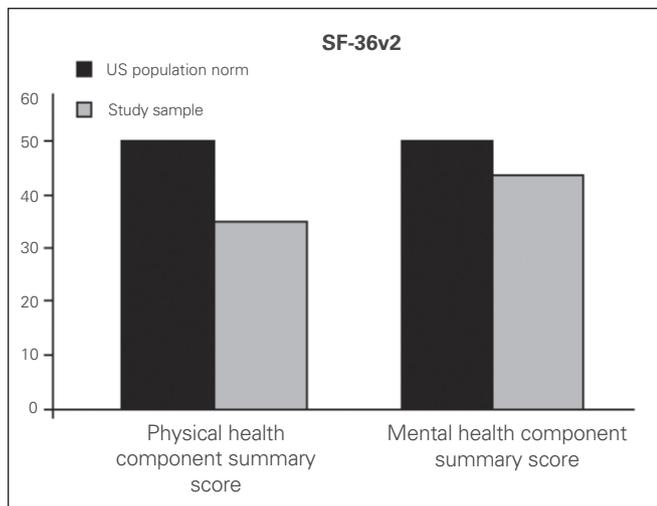


Fig. 1. Comparison of Short Form-36, version 2 (SF-36v2) scores from the study sample with the American population norms.

0.879 (95% CI 0.828–0.921; Table 4). The interrater reliability was 0.860 (95% CI 0.787–0.913) between the 2 orthopedic surgeons and the orthopedic fellow, 0.878 (95% CI 0.793–0.929) between the 2 residents and 0.893 (95% CI 0.819–0.938) between the 2 research coordinators.

Assessment of convergent validity

The correlations between the FIX-IT score and the SF-36v2 PCS score ranged from 0.682 to 0.770 (Table 5). The correlations between the FIX-IT score and the SF-36v2 physical functioning scale ranged from 0.681 to 0.758, and the correlation between the FIX-IT score and the SF-36v2 role-physical scale ranged from 0.677 to 0.786. The correlations between each procedure in the FIX-IT score and the SF-36v2 PCS are summarized in Table 5.

DISCUSSION

The use of a reliable, valid and responsive measure of fracture healing is essential for precisely estimating treatment effects in clinical trials.¹⁷ The FIX-IT measure is a recently developed, simple fracture healing assessment tool emphasizing outcomes that are likely important to patients. This preliminary study has demonstrated the FIX-IT measure has acceptable face and content validity and has shown that overall interrater reliability for the FIX-IT score among all 7 reviewers was 0.879 (95% CI 0.828–0.921), which demonstrates excellent agreement.¹⁴ The interrater reliability was above 80% among the 2 orthopedic surgeons and the orthopedic fellow, between the 2 residents,

Score, minimum: 0, maximum: 3*	Mean ± SD	Median	Min-max	Quartiles
Single-leg stand score	1.63 ± 1.19	2	0-3	25 1 50 2 75 3
Ambulation score	2.09 ± 0.93	2	0-3	25 2 50 2 75 3
Total score for ability to bear weight on fractured limb Minimum: 0, maximum: 6	3.72 ± 1.95	4	0-6	25 2 50 4 75 6
Palpation score	1.91 ± 0.75	2	1-3	25 1 50 2 75 2
Stress score	2.33 ± 0.80	3	0-3	25 2 50 3 75 3
Total score for pain at the fracture site Minimum: 0, maximum: 6	4.25 ± 1.35	4	1-6	25 3 50 4 75 5
Overall FIX-IT score Minimum: 0, maximum: 12	7.97 ± 2.73	8	1-12	25 6 50 8 75 10

FIX-IT = Function IndeX for Trauma; SD = standard deviation.
*Unless otherwise indicated.

and between the 2 research coordinators. This demonstrates that the FIX-IT measure has excellent reliability across different raters with different levels of clinical assessment skills and suggests that FIX-IT can be consistently administered by surgical trainees in clinical practices and in clinical studies by research coordinators.

Although the FIX-IT measure has adequate convergent validity with the SF-36v2, there are a couple of reasons that the correlation may not be perfect. First, generic health-related quality of life measures often lack sensitivity to detect smaller functional changes that may be affected by an orthopedic injury,² and it is possible that the FIX-IT measure better captured the patient's abilities than the SF-36v2. Also, the SF-36v2 elicits the patient's perspective on physical function whereas FIX-IT elicits the clinician's perspective on fracture healing. The expectation of healing may be different for the patient and the clinician, possibly impacting the ratings of their functioning.

Limitations

As the present study was a preliminary evaluation of the FIX-IT assessment, it had limitations. First, in the initial surgeon assessment of content, it may have been unclear to expert reviewers that the goal was not to further reduce the

items on the FIX-IT assessment. Surgeons may have felt that they had to indicate that at least something was not essential. Second, this was a convenience sample of patients from 1 surgeon's fracture clinic, limiting the generalizability of the findings. Third, the majority of the patients included in this study were assessed at least 12 months after the fracture, and many patients were being seen at the fracture clinic for complications. This is also evident in the patients' SF-36v2 scores, as they were lower than anticipated. The SPRINT study,¹⁹ a large randomized controlled trial evaluating reamed versus unreamed intramedullary nails, reported a physical component score of 42.9 ± 11 in the reamed group and 43.5 ± 11 in the unreamed group 1 year postinjury. This score is higher than the physical component score of 35.06 ± 9.77 found in the present study, indicating that patients in our study likely experienced more complications than the typical patient with a lower extremity fracture. Fourth, patients were only assessed once in our study, as opposed to being assessed over time to measure the progression of fracture healing, as in clinical practice or in a prospective clinical trial.

A strength of this study is that multiple raters with different clinical backgrounds and training levels independently assessed each included patient. The interrater agreement was acceptable among all raters, implying that the

Table 4. Interrater reliability of the FIX-IT measure

Item	Group; ICC (95% CI)			
	Overall, <i>n</i> = 7	Surgeons and orthopedic fellow, <i>n</i> = 3	Residents, <i>n</i> = 2	Research coordinators, <i>n</i> = 2
Single-leg stand score	0.834 (0.769–0.890)	0.811 (0.717–0.880)	0.825 (0.712–0.897)	0.805 (0.624–0.895)
Ambulation score	0.854 (0.795–0.904)	0.857 (0.784–0.911)	0.893 (0.819–0.938)	0.856 (0.760–0.916)
Total score, ability to bear weight on fractured limb	0.897 (0.852–0.933)	0.890 (0.831–0.932)	0.915 (0.855–0.951)	0.874 (0.763–0.931)
Palpation score	0.714 (0.620–0.803)	0.637 (0.494–0.759)	0.701 (0.529–0.818)	0.751 (0.599–0.851)
Stress score	0.725 (0.633–0.810)	0.685 (0.552–0.793)	0.737 (0.580–0.841)	0.771 (0.628–0.864)
Total score, pain at the fracture site	0.784 (0.705–0.854)	0.732 (0.612–0.827)	0.763 (0.618–0.858)	0.827 (0.716–0.898)
Total FIX-IT score	0.879 (0.828–0.921)	0.860 (0.787–0.913)	0.878 (0.793–0.929)	0.893 (0.819–0.938)

CI = confidence interval; FIX-IT = Function Index for Trauma; ICC = intraclass correlation coefficient.

Table 5. Correlation of the FIX-IT measure with the SF-36v2 and correlation of each procedure in the FIX-IT measure with the SF-36v2 — physical health component summary score

Item	Pearson correlation, values across individual assessors, range
Correlation of the FIX-IT measure with the SF-36v2	
SF-36v2 physical health component summary score (range)	0.682–0.770
SF-36v2 physical function scale score (range)	0.681–0.758
SF-36v2 role-physical scale score (range)	0.677–0.786
Correlation of each procedure in the FIX-IT measure with the SF-36v2 physical health component summary score (range)	
Single-leg stand score	0.499–0.594
Ambulation score	0.549–0.711
Palpation score	0.309–0.506
Stress score	0.444–0.503

FIX-IT = Function Index for Trauma; SF-36v2 = Short-form 36 version 2.

FIX-IT measure can be administered by study personnel or surgical trainees, reducing the demands of a clinical trial on the orthopedic surgeon.

CONCLUSION

The FIX-IT measure incorporates common clinical criteria into a simple assessment tool. The developers did not include questions about activities of daily living or return to work into the tool. Such questions were excluded in the FIX-IT measure because it was developed to be a simple index, and these questions are often included in other validated measures that are administered in patients participating in clinical trials.¹⁷ The developers also did not incorporate radiographic parameters into the assessment tool. As radiographic parameters are subjective, adjudication of these outcomes is becoming the gold standard,¹⁷ thus the developers opted to exclude radiographic outcomes from the FIX-IT measure.

Future research on the FIX-IT assessment should be conducted at multiple centres in larger numbers of patients, should include patients with fresh fractures and should measure the evaluation of fracture healing progression over time.

Acknowledgments: We thank the 5 orthopedic surgeons who assessed the FIX-IT index for face and content validity prior to the clinical study. We thank the reviewers for their support and commitment. We would also like to acknowledge Amgen Inc. for funding the study.

Competing interests: This study was funded by Amgen Inc. M. Bhandari is funded by a Canada Research Chair. S.M. Wasserman, N. Yurgin and R. Dent are employees of Amgen Inc and have stock options. No other competing interests declared. This study received research ethics board approval.

Contributors: M. Bhandari, S.M. Wasserman, N. Yurgin, S. Sprague and R.E. Dent designed the study. M. Bhandari, S.M. Wasserman, B. Petrisor and S. Sprague acquired the data, which M. Bhandari, S.M. Wasserman, N. Yurgin, S. Sprague and R.E. Dent analyzed. M. Bhandari, S.M. Wasserman and S. Sprague wrote the article. M. Bhandari, S.M. Wasserman, N. Yurgin, B. Petrisor and R.E. Dent reviewed the article. All authors approved its publication.

References

- Bhandari M, Guyatt GH, Swiontkowski MF, et al. A lack of consensus in the assessment of fracture healing among orthopaedic surgeons. *J Orthop Trauma* 2002;16:562-6.
- Morshed S, Corrales L, Genant H, et al. Outcome assessment in clinical trials of fracture-healing. *J Bone Joint Surg Am* 2008;90:62-7.
- Agel J, Swiontkowski MF. Guide to outcomes instruments for musculoskeletal trauma research. *J Orthop Trauma* 2006;20(8 Suppl):S1-146.
- Corrales LA, Morshed S, Bhandari M, et al. Variability in the assessment of fracture-healing in orthopaedic trauma studies. *J Bone Joint Surg Am* 2008;90:1862-8.
- Joslin CC, Eastaugh-Waring SJ, Hardy JR, et al. Weightbearing after tibial fracture as a guide to healing. *Clin Biomech (Bristol, Avon)* 2008;23:329-33.
- Aranzulla PJ, Muckle DS, Cunningham JL. A portable monitoring system for measuring weight-bearing during tibial fracture healing. *Med Eng Phys* 1998;20:543-8.
- Kershaw CJ, Cunningham JL, Kenwright J. Tibial external fixation, weight bearing, and fracture movement. *Clin Orthop Relat Res* 1993;293:28-36.
- Goodship AE, Kenwright J. The influence of induced micromovement upon the healing of experimental tibial fractures. *J Bone Joint Surg Br* 1985;67:650-5.
- Ware JE, Kosinski M, Turner-Bowker DM, et al. *How to score version 2 of the SF-12 Health Survey (with a supplement documenting version 1)*. Lincoln (RI): QualityMetric; 2002.
- Dogra AS, Ruiz AL, Marsh DR. Late outcome of isolated tibial fractures treated by intramedullary nailing: the correlation between disease-specific and generic outcome measures. *J Orthop Trauma* 2002;16:245-9.
- Zlowodzki M, Obremsky WT, Thomison JB, et al. Functional outcome after treatment of lower-extremity nonunions. *J Trauma* 2005;58:312-7.
- Bhandari M, Sprague S, Hanson B, et al. Health-related quality of life following operative treatment of unstable ankle fractures: a prospective observational study. *J Orthop Trauma* 2004;18:338-45.
- Bhandari M, Guyatt G, Tornetta P 3rd, et al.; Study to Prospectively Evaluate Reamed Intramedullary Nails in Patients with Tibial Fractures (SPRINT) Investigators. Randomized trial of reamed and unreamed intramedullary nailing of tibial shaft fractures. *J Bone Joint Surg Am* 2008;90:2567-78.
- Giraudeau B, Mary JY. Planning a reproducibility study: how many subjects and how many replicates per subject for an expected width of the 95 per cent confidence interval of the intraclass correlation coefficient. *Stat Med* 2001;20:3205-14.
- Fleiss JL. *Statistical methods for rates and proportions*. 2nd ed. New York: John Wiley & Sons; 1981.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74.
- Kooistra BW, Sprague S, Bhandari M, et al. Outcomes assessment in fracture healing trials: a primer. *J Orthop Trauma* 2010;24:S71-5.
- Butcher JL, MacKenzie EJ, Cushing B, et al. Long-term outcomes after lower extremity trauma. *J Trauma* 1996;41:4-9.
- SPRINT Investigators. Functional, general, and disability outcomes after tibial nailing: a randomized trial. Orthopaedic Trauma Association Annual Meeting; 2007 Oct. 17-20; Denver, Colo. Available: www.hwb.org/ota/am/ota07/otapa/OTA070422.htm (accessed 2010 Jan. 24).