BJR

TRAUMA

The radiographic union scale in tibial (RUST) fractures

RELIABILITY OF THE OUTCOME MEASURE AT AN INDEPENDENT CENTRE

J. M. Leow,
N. D. Clement,
T. Tawonsawatruk,
C. J. Simpson,
A. H. R. W. Simpson

University of Edinburgh, Edinburgh, United Kingdom

Objectives

The radiographic union score for tibial (RUST) fractures was developed by Whelan et al to assess the healing of tibial fractures following intramedullary nailing. In the current study, the repeatability and reliability of the RUST score was evaluated in an independent centre (a) using the original description, (b) after further interpretation of the description of the score, and (c) with the immediate post-operative radiograph available for comparison.

Methods

A total of 15 radiographs of tibial shaft fractures treated by intramedullary nailing (IM) were scored by three observers using the RUST system. Following discussion on how the criteria of the RUST system should be implemented, 45 sets (i.e. AP and lateral) of radiographs of IM nailed tibial fractures were scored by five observers. Finally, these 45 sets of radiographs were rescored with the baseline post-operative radiograph available for comparison.

Results

The initial intraclass correlation (ICC) on the first 15 sets of radiographs was 0.67 (95% CI 0.63 to 0.71). However, the original description was being interpreted in different ways. After agreeing on the interpretation, the ICC on the second cohort improved to 0.75. The ICC improved even further to 0.79, when the baseline post-operative radiographs were available for comparison.

Conclusion

This study demonstrates that the RUST scoring system is a reliable and repeatable outcome measure for assessing tibial fracture healing. Further improvement in the reliability of the scoring system can be obtained if the radiographs are compared with the baseline post-operative radiographs.

Cite this article: *Bone Joint Res* 2016;5:116–121.

Keywords: score; interobserver reliability; tibial fractures; radiographic union; intramedullary fixation

■ J. M. Leow, Medical Student, College of Medicine and Veterinary Medicine, University of Edinburgh, 11 Montague Street, Edinburgh EH8 9QT, UK.
■ T. Tawonsawatruk, MD, PhD, Speciality Trainee, Trauma and Orthopaedics, Orthopaedic Department, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand.
■ C. J. Simpson, Medical Student, School of Medicine and Dentistry, University of Aberdeen, Westburn Dr, Aberdeen AB25 3BZ, UK.
■ N. D. Clement, MRCS, PhD, Speciality Trainee, Trauma and Orthopaedics.
■ A. H. R. W. Simpson, MA (Cantab), DM (Oxon), FRCS (England & Edinburgh), George Harrison Law Professor of Orthopaedic Surgery, Department of Orthopaedics, Royal Infirmary of Edinburgh, Little France Crescent, Edinburgh EH16 4SA, UK.

Correspondence should be sent to Mr J.M. Leow; e-mail: leowjunmin@gmail.com

doi:10.1302/2046-3758.54.2000628

Bone Joint Res 2016;5:116–121.
Received: 20 November 2015;
Accepted: 26 February 2016

Article focus

- To evaluate the reliability and repeatability of the radiographic union scale in tibial (RUST) fracture score, and whether it improves when radiographs that were scored were compared with the baseline post-operative radiographs.

Key messages

- RUST is a reliable outcome measure which can be repeated for assessing tibial fracture healing.
- The reliability of the scoring system is increased by agreeing the interpretation of the scoring system and by comparing the radiographs being scored to the baseline post-operative radiographs.

Strengths and limitations

- The radiographs were not standardised, but they were the views obtained for standard clinical care making the results generalisable.
- Thus, the RUST score has the potential to serve as a reliable scoring system to help quantify healing in both clinical and research settings.

Introduction

While treatment of tibial fractures by intramedullary nailing has been shown to have good outcomes, there remains a lack of consensus among orthopaedic surgeons in the assessment of bony union.[1] In addition, the radiographic definition of delayed union is also vague with varying criteria depending on the preference of the assessor. A review of the reliability and validity of radiographic assessments of tibial fractures highlighted the need for an "accurate assessment of radiographic healing".[2]

Radiographic cortical bridging by callus and the lack of a fracture line offer the most reliable signs of bone healing between observers.[3] These findings led to the development of the radiographic union scale in tibial fractures (RUST) score by Whelan et al,[4] which uses these radiographic signs to assess healing.

The RUST score is a novel fracture assessment tool that was developed to help standardise the radiographic assessment of tibial fractures.[4] This score assesses cortical bridging, which has been shown to correlate with the biomechanical strength of the fracture site in *in vivo* models.[5] In the original paper by Whelan et al,[4] the authors demonstrated that the RUST score is a reliable assessment tool of fracture healing with good agreement among five observers (intraclass correlation of 0.86). The RUST score has been used in various clinical studies since its introduction and further validated for use in small animal models.[6] However, we are unaware of any independent validation and/or reliability study for the RUST score, aside from the original study. In addition, the effect of serial radiograph assessment upon observer agreement has not been tested. In particular, the effect of an observer being able to compare the immediate post-operative film with subsequent radiographs (in order to take into account the fracture configuration) has not been evaluated. Further confirmation of the reliability of the RUST score may substantiate the use of the score for both clinical assessment and as a research tool.

The primary aim of this study was to assess the reliability of the RUST score in an independent centre when using the same methods described by the original authors. Our secondary aim was to assess the interobserver agreement of the RUST score with the addition of the baseline post-operative radiographs for comparison.

Patients and Methods

Ethical approval was obtained for this study by the local ethics committee. Patients were retrospectively identified from a radiological electronic database. Patients aged 16 years or over with tibial shaft fractures, also classified as type 42 fractures by the AO Foundation,[7] treated at the study centre between July 2007 and June 2013 were included in the study. All patients from the catchment area who were treated in the study centre, but resided outside these areas, were excluded from analysis. Conversely, all patients receiving their initial management outside our catchment area but who resided within it were included. This resulted in 393 fractures being available for study, of which 264 were managed with a reamed locked intramedullary nail. In addition to the immediate post-operative films, standard radiographic follow-up assessment included anteroposterior (AP) and lateral radiographs at six weeks, three and six months. At this stage, the patient was either discharged if the fracture was united, or subjected to further radiographs or reoperation if the fracture was not united.

Table I. Overview of radiographic union scale in tibial fracture (RUST).

Score per cortex	Callus	Fracture Line
1	Absent	Visible
2	Present	Visible
3	Present	Invisible

A pilot study was carried out to standardise the interpretation of the score as opposed to the use of the score. This was carried out on 15 radiographs by three observers. The observers were given the description of the RUST score as in the original paper by Whelan et al.[4] The intraclass correlation coefficient (ICC) was then calculated. The interpretation of the grading of the scores (see below) were then discussed among the three observers.

After a consensus was achieved regarding the scoring, the second part of the study was then performed: 45 sets of anteroposterior and lateral radiographs of tibial diaphyseal fractures treated with intramedullary nailing[4] were randomly selected from the cohort of patients by an independent researcher. Radiographs with visible staples or casts were excluded as this may have provided hints for the fracture age. All radiographs chosen were at least two weeks from the date of surgery. A group of five reviewers, which included three orthopaedic surgeons and two independent researchers, assigned RUST scores to each of the 45 sets of radiographs. To reduce bias, the patients' details, history, and fracture age were blinded from the reviewers. Each tibial cortex (anterior, posterior, medial and lateral) was assigned a RUST score of 1 to 3, based on the appearance. A cortex with a visible fracture line and no callus was given a score of 1, a cortex where callus and a visible fracture line was present was scored as 2, and a cortex with bridging callus and no fracture line within the callus bridge was scored as 3 (Table I). The scores of all cortices were then combined to give a minimum score of 4 (definitely not healed) and a maximum of 12 (completely healed).

The reviewers assessed the radiographs independently of one another and assigned a RUST score to each anonymised patient. They were asked to review the same radiographs again four weeks later in a different numerical order, and assigned a RUST score once again to assess for intra-observer variation. In addition, at eight weeks, the reviewers were then asked to assign a RUST score with the immediate post-operative radiograph available

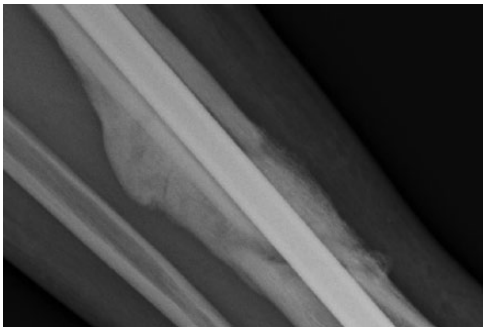

Fig. 1a

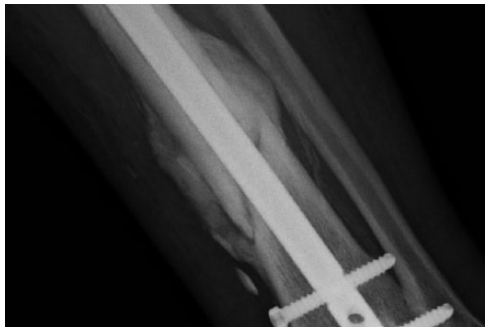

Fig. 1b

Anteroposterior radiographs a) and b) of the tibia and fibula which show bridging callus, yet fracture lines can still be seen between the original cortices.

for comparison and evaluate any improvement in the interobserver reliability of RUST.

Statistical analysis. The intraclass correlation coefficient (ICC) with 95% confidence intervals was used to quantify agreement of the RUST score, a continuous variable, between reviewers. The interobserver reliability of the five examiners was then calculated using SPSS Version 18.0 (SPSS, Chicago, Illinois). Results from ICC model 2 was selected as this model is used when the same methods or raters perform the evaluations in all cases. ICC is interpreted as follows: 0 to 0.2 indicates *poor* agreement: 0.3 to 0.4 indicates *fair* agreement; 0.5 to 0.6 indicates *moderate* agreement; 0.7 to 0.8 indicates *strong* agreement; and > 0.8 indicates *almost perfect* agreement.[8]

Results

In the pilot study, the interobserver ICC was 0.67 (95% CI 0.63 to 0.71). It was noted that the poor concordance was particularly evident on a subset of individual cortices scored 2 or 3. On closer scrutiny, it became clear that the description in the original paper on whether a fracture line was still present was being interpreted in two different ways: either there was no discontinuity/fracture in the bridging callus or there was no discontinuity/fracture in the tibial cortex. In these contentious radiographs there were 'cortices' that still had a visible fracture line in the cortex, but had a continuous bridging callus without a fracture line within the callus (Figs 1 and 2). It was determined that the fracture line should only apply to the callus, and therefore, a callus which had a discontinuity/fracture line was scored as 2 (Fig. 2b), while a callus which was bridged (i.e. no discontinuity/fracture line) was scored as 3 (Fig. 2b). This interpretation was then used for the remainder of the study.

In the second part of the study, the RUST score of all 45 radiographs was measured. The values of the RUST score ranged from 4 to 12, with a mean score of 9.1 (standard deviation (SD) 2.4) (Fig. 3). The agreement between five scorers was strong at 0.75 (95% CIs 0.65 to 0.84) for each individual score. However, when the average total measure of the RUST score was used, the ICC increased to 0.94 (95% CI 0.90 to 0.96), showing almost perfect agreement. The mean ICC for the intra-observer variability was 0.79 (95% CI 0.66 to 0.86).

The inter-observer ICC increased from 0.75 to 0.79 (95% CI 0.71 to 0.87) when the post-operative radiograph was available. When the average measure of all scorers was used, agreement did not increase significantly with the availability of post-operative radiographs (ICC 0.94, 95% CI 0.90 to 0.96 and ICC 0.95, 95% CI 0.92 to 0.97). The ICC of three orthopaedic surgeons was calculated to be 0.78 (95% CI 0.67 to 0.86), increasing to 0.86 (95% CI 0.78 to 0.91) with the use of post-operative radiographs. The two independent researchers had an ICC of 0.71 (95% CI 0.53 to 0.83), which was unchanged (ICC 0.71; 95% CI 0.52 to 0.83) despite the use of post-operative radiographs.

Discussion

This study supports the reliability of the RUST score in grading the healing of tibial diaphyseal fractures, demonstrating strong inter- and intra-observer agreement. In addition, we have also demonstrated that for individual component measures, the reliability improves if the post-operative radiograph is available for comparison. However, the overall mean total score was not significantly influenced by knowledge of the immediate post-operative fracture pattern.

There are several limitations to this study. First, it is important to note, as highlighted by the original authors of the RUST score,[4] that the inter- and intra-observer reliability analyses only the precision of a score or classification system, and not the accuracy. However, the results here do support the use of the RUST to standardise outcomes when comparing different investigations of tibial fractures managed with intramedullary fixation. Secondly, there were a limited number of observers used in the development of the RUST score, which would require further verification with a large sample size analysed by varying levels of experience of the observers. Thirdly, it is unknown whether the RUST system can distinguish a healed fracture from a nonunion, because no comparison was made to other patient ratings, biomechanical strength, or pain scores. Despite these limitations, the

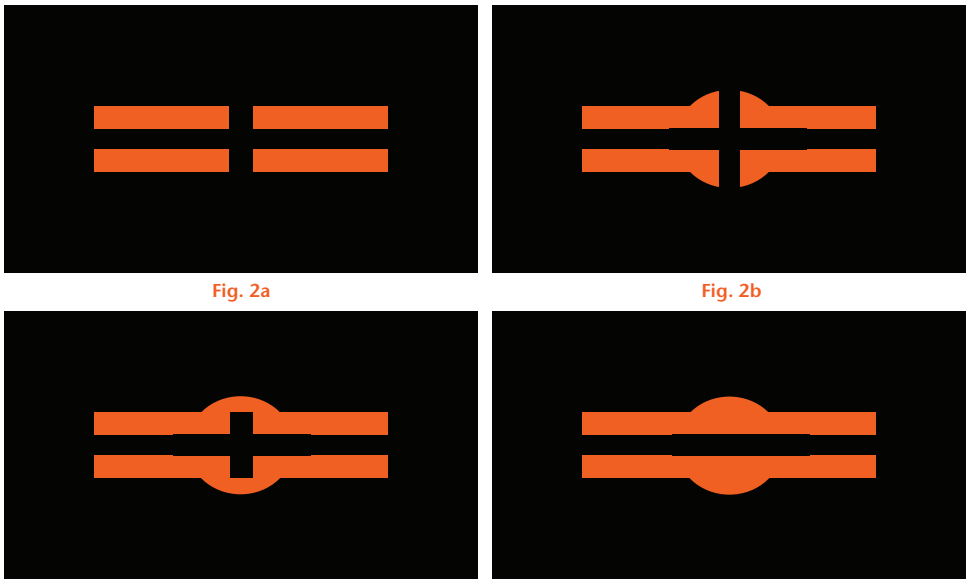

Fig. 2a Fig. 2b

Fig. 2c Fig. 2d

Diagrams showing a) a fracture with a fracture line and no callus formation; this would be assigned a radiographic union scale in tibial (RUST) fracture score of 1; b) a fracture with callus formation and a fracture line; this is scored as 2; c) a fracture with bridging callus, but the fracture line is still visible across both cortices; this is scored as 3 and d) complete bridging of the callus with no evidence of fracture line and is scored as 3.

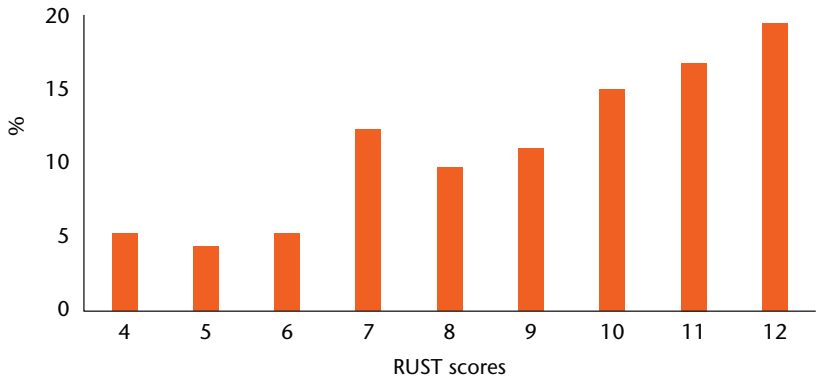

Fig. 3

Graph showing the percentage of radiographic union scale in tibial (RUST) fractures scores of radiographs

RUST score may function as a supplemental tool that clinicians can use to assess tibial fracture healing. In addition, it has the potential to serve as a reliable scoring system to help quantify healing in research settings.

Classically, Sarmiento et al[9] defined union of tibial diaphyseal fractures managed in a functional brace as: no pain on weight bearing, no movement at the fracture site and callus evident on radiographic assessment. Weight-bearing status has been shown to correlate relatively well with fracture stiffness in tibial fractures treated with external fixation,[10] however, the surgeon's ability to judge stiffness and weight-bearing capability based on physical exam alone is not reliable. Webb et al[11] demonstrated that manual assessment of stiffness by orthopaedic surgeons was not superior to that by medical students. Additionally, it has been shown that physicians, regardless of number of years of experience, are not consistent when assessing the increasing stiffness of fractures with time.[12] Pain on palpation at the site of injury is also currently widely used among physicians to judge union, however, it is a highly subjective outcome given individual and cultural differences in perception and tolerance level of pain among the patient population. In addition, the assessment of a fracture with a reamed nail (which provides stability to allow early weight bearing) results in difficulty in the clinical assessment of fracture union. With the knowledge that a definition is essential for fracture union and the assessment of healing, a quantitative method is required for both clinical and research purposes.

Plain radiographs, radionuclide imaging, computed tomography (CT), ultrasonography and resonance frequency analysis have yielded good results in defining

fracture consolidation.[13] CT scans have been reported to have 100% sensitivity for detecting nonunion but are limited by a low specificity of 62%.[14] However, such interventions in routine clinical practice are expensive and expose the patients to potentially harmful radiation. Alternatively, ultrasound assessment does not expose the patient to radiation but does not penetrate cortical bone. Despite this limitation, there is evidence that ultrasound is able to detect callus formation before radiographic changes are visible.[15,16] Ultrasound has additional advantages over other imaging modalities including lower cost, no ionising radiation exposure, and the fact that it is non-invasive. However, its use and interpretation of findings are thought to be highly dependent on the operator's expertise. Hence, the most convenient method of assessing fracture healing is currently by radiographic assessment, which is already in place as part of the patients' routine follow-up. The RUST score would seem to be the most reliable method to assess bone healing on a routine basis.

When scoring the radiographs, overlapping bone could be mistaken as callus, for example in a spiral fracture with a minor degree of displacement (Fig. 4). In contrast, rigid fixation of the fracture may obscure the fracture line, giving the false appearance of a fully remodelled fracture (Fig. 5). This was the rationale for scoring radiographs by comparing them with the immediate post-operative radiographs. From our results, we found that while the ICC does increase slightly from 0.75 to 0.79 when the post-operative radiograph is available, this was not a significant increase. However, the authors considered that it made the process of assigning RUST scores easier, and represented how fracture assessment was carried out in a clinical setting by comparing the current radiograph with post-operative films.

In the original paper, Whelan et al[4] defined a RUST score of 3 as a fracture with bridging callus with no evidence of a fracture line. In the initial pilot stage of our study, we found that this could be interpreted in two ways - either as an absence of fracture line in the original cortex or in the bridging callus. The reliability of the RUST score was improved by agreeing through discussion that a fracture with callus, which has a fracture line/discontinuity within, is scored as 2 (Fig. 2b) while a bridging callus with no discontinuity is scored as 3 (Fig. 2c).

Having demonstrated the reliability of this score, future research can focus on applying RUST in a clinical setting to investigate whether it can be used to distinguish normal healing fractures from nonunions, at an early stage.

In conclusion, this work has demonstrated that the RUST score has strong intra- and inter-observer agreement, provided that the interpretation of the grade 2 or grade 3 scores is clarified, and is a reliable and repeatable outcome measure for assessing tibial fracture healing. A small further improvement in the reliability of the scoring system can be made if the radiographs which are scored can be compared with the baseline post-operative radiographs.

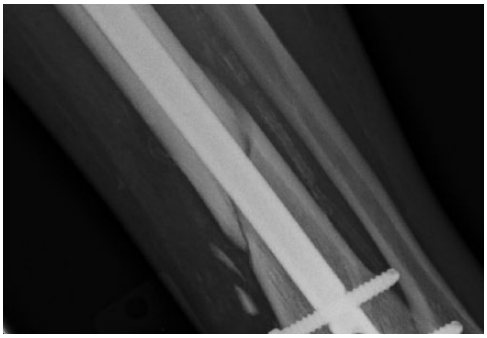

Fig. 4

This anteroposterior radiograph of the tibia and fibula may seem to have formation of callus around the lateral cortex despite being taken immediately post-operation. The appearance of bridging bone is due to the spiral pattern and slight displacement, where in fact no callus has yet to form.

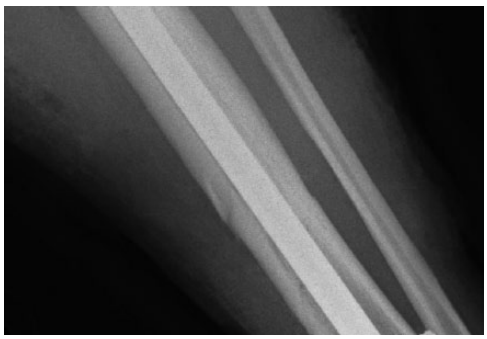

Fig. 5

Anteroposterior radiograph showing the tibia and fibula. It was taken immediately after operation and shows barely visible fracture lines, but no callus.

References

1. **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;16:562-566.
2. **Kooistra BW, Dijkman BG, Busse JW, et al.** The radiographic union scale in tibial fractures: reliability and validity. *J Orthop Trauma* 2010;24:S81-S86.
3. **Whelan DB, Bhandari M, McKee MD, et al.** Interobserver and intraobserver variation in the assessment of the healing of tibial fractures after intramedullary fixation. *J Bone Joint Surg [Br]* 2002;84-B:15-18.
4. **Whelan DB, Bhandari M, Stephen D, et al.** Development of the radiographic union score for tibial fractures for the assessment of tibial fracture healing after intramedullary fixation. *J Trauma* 2010;68:629-632.
5. **Panjabi MM, Walter SD, Karuda M, White AA, Lawson JP.** Correlations of radiographic analysis of healing fractures with strength: a statistical analysis of experimental osteotomies. *J Orthop Res* 1985;3:212-218.
6. **Tawonsawatruk T, Hamilton DF, Simpson AH.** Validation of the use of radiographic fracture-healing scores in a small animal model. *J Orthop Res* 2014;32:1117-1119.
7. **Theerachai Apivatthakakul, Surapong Anuraklekha, George Babikian, et al.** Tibial shaft – Diagnosis – AO Surgery Reference. Available through URL: https://www2.aofoundation.org/wps/portal/surgery?showPage=diagnosis&bone=Tibia&segment=Shaft

8. **Landis JR, Koch GG.** The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159-174.
9. **Sarmiento A, Sobol PA, Sew Hoy AL, et al.** Prefabricated functional braces for the treatment of fractures of the tibial diaphysis. *J Bone Joint Surg [Am]* 1984;66-A:1328-1339.
10. **Joslin CC, Eastaugh-Waring SJ, Hardy JR, Cunningham JL.** Weight bearing after tibial fracture as a guide to healing. *Clin Biomech (Bristol, Avon)* 2008;23:329-333.
11. **Webb J, Herling G, Gardner T, Kenwright J, Simpson AH.** Manual assessment of fracture stiffness. *Injury* 1996;27:319-320.
12. **Hammer R, Norrbom H.** Evaluation of fracture stability. A mechanical simulator for assessment of clinical judgement. *Acta Orthop Scand* 1984;55:330-333.
13. **Morshed S.** Current Options for Determining Fracture Union. *Adv Med* 2014;2014:708574.
14. **Bhattacharyya T, Bouchard KA, Phadke A, et al.** The accuracy of computed tomography for the diagnosis of tibial nonunion. *J Bone Joint Surg [Am]* 2006;88-A:692-697.
15. **Craig JG, Jacobson JA, Moed BR.** Ultrasound of fracture and bone healing. *Radiol Clin North Am* 1999;37:737-751, ix.
16. **Moed BR, Subramanian S, van Holsbeeck M, et al.** Ultrasound for the early diagnosis of tibial fracture healing after static interlocked nailing without reaming: clinical results. *J Orthop Trauma* 1998;12:206-213.

Funding Statement
- Departmental funding has been received from Stryker, RCSEd, OTCF, ARUK, ORUK, and EPSRC, none of which is related to this article.

Author Contribution
- J. M. Leow: Data collection, Data analysis, Writing the paper.
- N. D. Clement: Data collection, Data analysis, Writing the paper.
- T. Tawonsawatruk: Study design, Data collection, Data analysis, Reviewing the paper.
- C. J. Simpson: Data collection, Data analysis, Reviewing the paper.
- A. H. R. W. Simpson: Study design, Data collection, Data analysis, Writing the paper.

ICMJE conflict of interest
- None declared.