A commentary by Gregory P. Guyton, MD, is linked to the online version of this article.

Concomitant Unstable and Stable Gravity Stress Tests on Weight-Bearing Stable Weber B Ankle Fractures Treated Nonoperatively

A 2-Year Outcome Study

Martin G. Gregersen, PT, MSc, Hilde Stendal Robinson, PT, PhD, and Marius Molund, MD, PhD

Investigation performed at the Østfold Hospital Trust, Grålum, Norway

Background: Replacing gravity stress tests with weight-bearing radiographs to evaluate the stability of Weber B (also called Lauge-Hansen supination-external rotation [SER]) ankle fractures results in a lower surgery rate, thus avoiding associated risks and complications. Still, nonoperative treatment of weight-bearing stable fractures is controversial because of the scarcity of strong evidence. We investigated the influence of a concomitant unstable gravity stress test compared with a stable gravity stress test on outcomes after nonoperative treatment of weight-bearing stable fractures.

Methods: We performed a prospective, noninferiority study on 149 patients with Weber B ankle fractures and stable weight-bearing radiographs. Gravity stress radiographs classified fractures as stable (SER2 [n = 88]) or partially unstable (SER4a [n = 61]). All were treated with a functional orthosis and weight-bearing was allowed; patients were followed for 2 years. The primary outcome was the Manchester-Oxford Foot and Ankle Questionnaire (MOXFQ), with a range from 0 to 100, in which lower scores indicate fewer symptoms. A noninferiority margin was prospectively defined as 7.5 points. The secondary outcomes included the Olerud-Molander Ankle Score, assessment of ankle congruence, and treatment-related adverse events.

Results: The primary outcome data were available for 144 (96.6%) of 149 participants at 2 years. The between-group difference in the MOXFQ score was 1.0 point (95% confidence interval, −1.4 to 3.4 points; p = 0.397) in favor of the SER2 group, consistent with noninferiority. We found no appreciable between-group differences for any other outcome.

Conclusions: In Weber B/SER ankle fractures that are stable on weight-bearing radiographs, are treated with removable orthoses, and are allowed to bear weight, a concomitant unstable gravity stress test (SER4a) was not associated with worse patient-reported or radiographic outcomes compared with a stable gravity stress test (SER2) at the 2-year follow-up. Thus, the identification of stress instability seems redundant, which questions the applicability of stress instability for surgical decision-making.

Level of Evidence: Diagnostic Level II. See Instructions for Authors for a complete description of levels of evidence.

Gravity stress tests and weight-bearing radiographs are alternative methods for evaluating deltoid ligament integrity in Weber B, Lauge-Hansen, supination-external rotation (SER), or OTA/AO 44B[1] ankle fractures, thus determining the choice between a surgical procedure or nonoperative treatment. Both are based on assessing the alignment of the ankle mortise. If the talus is anatomically aligned with the tibia, the ankle is deemed stable. Instability is indicated in 35% to 48% of ankle fractures when gravity stress tests are used[2-5], but in only 1% to 3% of fractures when weight-bearing radiographs are used[2-4], meaning that the ankle mortise often changes from nonanatomic to anatomic (i.e., an apparently unstable fracture to an apparently stable

Disclosure: The **Disclosure of Potential Conflicts of Interest** forms are provided with the online version of the article (http://links.lww.com/JBJS/H632).

A **data-sharing statement** is provided with the online version of the article (http://links.lww.com/JBJS/H682).

one) after the orientation of the leg is shifted and weight is applied to it. This is rationalized by a possible partial rupture of the deltoid ligament, with the deep posterior portion of the ligament remaining intact[3,6,7]. Some authors have suggested classifying such an injury as SER4a[3,6]. When both tests are stable, probably because there is no deltoid ligament injury, the appropriate classification is SER2. With both tests deemed unstable, the proper classification is SER4b[3,6], which theoretically results from complete deltoid ligament disruption.

Stress-stable (SER2) fractures are effectively treated with functional orthoses[5,8], as casting or restricted weight-bearing is not necessary[5]. However, evidence for treating fractures with an unstable stress test (corresponding to SER4a-b) has been limited and inconclusive. A solitary randomized controlled trial (RCT) found no outcome differences between operative and nonoperative treatment[9]. Cohorts have suggested that fractures deemed to be stable on weight-bearing radiographs (corresponding to SER2 and SER4a) can be managed nonoperatively with satisfactory results[2-4,10-13]. These studies were small and retrospective[4,10-13], and treatment protocols varied[4]; thus, the treatment choice for SER4a fractures remains controversial.

In this prospective, noninferiority study on weight-bearing, stable, Weber B fractures treated nonoperatively, we determined the effect of a concomitant unstable gravity stress test (SER4a) compared with a stable gravity stress test (SER2) on patient-reported outcomes and adverse events. The rationale for the noninferiority design was that a surgical procedure for SER4a fractures could be regarded as overtreatment, and could expose patients to an unnecessary risk of surgical complications, if SER4a fractures achieve clinical outcomes that are not unacceptably worse than those of SER2 fractures after both received identical nonoperative treatment (i.e., the evaluation compared a new therapy for SER4a fractures with the established treatment for SER2 fractures). We hypothesized that there would be no outcome difference between ankles with gravity stress test results indicating a stable or unstable fracture if the weight-bearing radiographs indicate stability.

Materials and Methods

Study Design

This study was conducted at the Østfold Hospital Trust, Grålum, Norway, from January 2019 to May 2020. The 2-year follow-up was concluded in June 2022. The study was registered at ClinicalTrials.gov (NCT03831009) and reporting followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) checklist[14] and the Consolidated Standards of Reporting Trials (CONSORT) statement for noninferiority trials[15]. The Regional Committee for Medical and Health Research approved the study (24994). Study participants gave informed consent before taking part. A previous case-control study was published on short-term radiographic outcomes using the same original patient sample[3].

Participants

We evaluated patients who were 18 to 80 years of age and had an isolated Weber B/Lauge-Hansen SER2 or SER4a ankle fracture and a medial clear space (MCS) measurement of <7.0 mm on non-weight-bearing radiographs for eligibility. We excluded patients with pre-hospital fracture reduction, previous ipsilateral or contralateral ankle fracture, bimalleolar or trimalleolar fractures, other osseous injury, major ankle or hindfoot surgery, pre-injury walking aid dependence, peripheral neuropathy, presumed inability to comply with the treatment protocol, or no fluency in Norwegian.

Procedure

Stability Assessment

After identifying a Weber B/SER fracture with <7.0-mm MCS, a below-the-knee plaster cast was applied with the ankle held at 90° and maintained until final stability examinations. Weight-bearing was not allowed in the cast. We delayed stability examinations by 3 to 7 days after the fracture, anticipating pain reduction. Investigators performed stability examinations in the outpatient clinic, evaluating bilateral weight-bearing and gravity stress radiographs. This was according to the standard procedure at our hospital, and we invited patients to participate in the study at the time of stability evaluation. Weight-bearing radiographs were obtained without the cast. The foot had to be plantigrade-loaded with >50% of body weight. Using a bathroom scale, we verified patients' ability to perform weight-bearing before radiography. Gravity stress tests[16] were performed with the patient positioned on the side, lying on a flat examination table, with the medial side of the injured ankle facing up. The distal half of the injured leg was placed outside the table's edge, and the patient was instructed to relax. A radiopaque sphere (32-mm diameter) was taped to the patient's ankle for calibration. A Sectra Picture Archiving and Communications System (Sectra AB) was used for radiograph measurements. The size of the MCS was considered an expression of tibiotalar congruence and an indirect measure of ankle stability. The MCS was recorded in millimeters on mortise view radiographs. Consistent with previous descriptions[3,17], we defined the MCS as the distance between the medial border of the talus and the lateral border of the medial malleolus on a line parallel to and 5.0 mm below the talar dome, which has demonstrated excellent reliability[3]. The ankle mortise was considered stable if the MCS measurement was <5.0 mm. Correspondingly, it was considered unstable if the MCS measurement was ≥5.0 mm and had a ≥1.0-mm increase compared with the uninjured ankle. These criteria were used in interpreting weight-bearing and gravity stress radiographs. Orthopaedic residents or physiotherapists measured the radiographs and were not blinded to clinical or outcome data.

Group Assignment

Participants with weight-bearing radiographs deemed stable (corresponding to SER2 and SER4a fracture types) underwent subsequent gravity stress radiographs. They were then assigned to the SER2 group if the gravity stress test result was considered stable or the SER4a group if deemed unstable. Participants with unstable weight-bearing radiographs were classified as having SER4b injuries, were treated operatively, and were excluded. Participants were unaware of their assignments, but clinicians were informed.

Interventions

Both groups were treated with functional orthoses (Aircast and Air-Stirrup; DJO Global) for 6 weeks. Weight-bearing on the injured ankle as tolerated was allowed. We advised patients to wear the orthosis as needed and always when walking. Temporary removal to perform range-of-motion exercises was encouraged.

Follow-up

Investigators conducted follow-up visits at 2, 6, and 12 weeks and 1 and 2 years after the fracture. Visits included clinical examinations, ankle weight-bearing radiographs, and completion of questionnaires.

Primary Outcome

The primary outcome was the Manchester-Oxford Foot and Ankle Questionnaire (MOXFQ), a 16-question, patient-reported questionnaire used to assess symptoms and quality of life in patients with ankle pathology[18-20]. We used the metric MOXFQ index (0 to 100, in which lower scores indicate fewer symptoms)[21]. When planning this study, no minimal clinically important difference (MCID) value was available for the MOXFQ index score, despite existing estimates for its subscores[22,23]. The minimal detectable change was defined as 5.0 points[24].

Secondary Outcomes

Secondary outcomes included the Olerud-Molander Ankle Score (OMAS), which is patient-reported and ranges from 0 to 100 (higher scores indicate fewer symptoms). The OMAS has 9 items focusing on symptoms and ankle function[25] and has been validated in Norwegian[26]. The MCID for patients with ankle fractures was previously defined as 8.8 points[5] and later as 9.5 points with a minimal detectable change of 4.7 points[4]. We also collected the outcomes of the American Foot & Ankle Society (AOFAS) ankle-hindfoot score, a numeric rating scale of pain, the EuroQol 5 Dimensions (EQ-5D), bilateral ankle range of motion, and bilateral calf circumference. These outcomes were not analyzed or reported in this study. However, the data are available upon request.

Radiographic Outcomes

We assessed weight-bearing radiographic congruity of injured ankles compared with normal ankles at 2 years by evaluating differences in MCS measurements of the injured ankle compared with the uninjured ankle. The MCS was recorded in millimeters on mortise view radiographs. We defined the MCS as the distance between the medial border of the talus and the lateral border of the medial malleolus on a line parallel to and 5.0 mm below the talar dome. Orthopaedic residents or physiotherapists measured the radiographs and they were not blinded to treatment assignment.

Treatment-Related Adverse Events

The incidence of treatment-related adverse events and thromboembolic events was monitored until study closure. Fracture healing was determined by callus formation on radiographs and the absence of pain at the fracture site. Delayed union was registered if a union had not occurred within 12 weeks after the fracture, and nonunion was registered if a union had not occurred within 6 months after the fracture.

Demographic Characteristics and Covariates

Sex, age, body mass index, and smoking status were recorded at baseline.

Statistical Analysis

We calculated that a sample size of 126 would be sufficient to ensure 80% power at a 5% significance level to detect a difference of 7.5 points in the MOXFQ. To compensate for the patients lost to follow-up, we enrolled 149 patients.

We decided on 7.5 points as the noninferiority limit. This was larger than the minimal detectable change, but lower than previously reported MCID values for the subscores. We contend that this was a conservative approach, thus increasing sensitivity to detect outcome differences and reducing the risk of falsely claiming noninferiority. For the OMAS, 8.8 points constituted the noninferiority limit. Tibiotalar congruity on radiographs, comparing injured ankles with normal ankles after 2 years, was evaluated with an equivalence analysis using a predefined margin of 1.0 mm, consistent with our previous study[3]. The remaining secondary outcomes were assessed using superiority analyses to determine differences.

Descriptive data were presented as means with standard deviations, medians and ranges, or frequencies and percentages. Because patient-reported data did not satisfy normality assumptions, we performed a bootstrap procedure (3,000 repetitions, seed = 2,345) on the mean MOXFQ and OMAS difference for each time point. The 2-year between-group differences, with 2-sided 95% confidence intervals (CIs), were derived from the bootstrap procedure. If the upper limit of the 95% CI for the difference between the groups was below the noninferiority margin, noninferiority could be claimed. Student t tests were used to evaluate the difference in radiographic MCS of the injured ankle at 2 years compared with the MCS of the uninjured ankle at baseline. The patient's age and body mass index were prospectively identified as potential confounders for patient-reported ankle function; diabetes was not considered in the same way, as patients with peripheral neuropathy were excluded. Simple linear and multiple regression were used to control and adjust for confounding.

We performed both intention-to-treat and as-treated analyses. Moreover, we performed post hoc sensitivity analyses to evaluate if alternative noninferiority margins led to different interpretations of between-group differences in patient-reported outcome data. We used the above-referenced MCID values from contemporary studies and minimal detectable changes as the alternative margins. Statistical analyses were performed using Stata (release 17; StataCorp).

Source of Funding

This study was funded by the Østfold Hospital Trust and the South-Eastern Norway Regional Health Authority (grant 2023014).

Results

We screened 167 patients with a Weber B/SER ankle fracture and <7.0-mm MCS on initial radiographs for eligibility: 153 were eligible, and 151 agreed to participate in the study. We excluded the other 16 patients. Weight-bearing radiographs were made for all patients within 7 days after the fracture. Among the 151 consenting participants, 2 (1.3%) had weight-bearing radiographs considered unstable (classified SER4b) and underwent operative treatment. The remaining 149 participants (98.7%) had stable weight-bearing radiographs (corresponding to SER2 or SER4a fracture types) and were included in the study. Of these, 88 (59.1%) were SER2 injuries, and 61 (40.9%) were SER4a injuries according to the interpretation of the gravity stress result. Of the 149 included patients, 147 (98.7%) had follow-up at 6 weeks; 141 (94.6%), at 12 weeks; 104 (69.8%), at 1 year; and 144 (96.6%), at 2 years. Figure 1 is a flow diagram for the study.

The mean age (and standard deviation) was 50 ± 15 years, and 89 (59.7%) were female (see Table I).

Patient-Reported Outcomes

The primary outcome scores of SER4a fractures at 2 years were noninferior to those of SER2 fractures (mean difference in the MOXFQ score, 1.0 point [95% CI, −1.4 to 3.4 points]; p = 0.397), with the SER4a group reporting a mean MOXFQ score of 3 ± 9 and the SER2 groups reporting a mean score of 2 ± 4 (see Table II).

Figure 2 illustrates the interpretation of noninferiority of MOXFQ and OMAS scores over the study course using 95% CIs and noninferiority margins. Appendix Figures S1 and S2 show treatment responses for the MOXFQ and OMAS in the groups over the course of the study. Conclusions about noninferiority at 2 years remained robust in sensitivity analyses.

None of the predefined potential confounders significantly predicted MOXFQ scores in the multivariable regression models (see Appendix Table S1).

Radiographic Outcomes

Ankle radiographs made at 2-year follow-up were available in 123 participants (82.6%). The mean difference in the

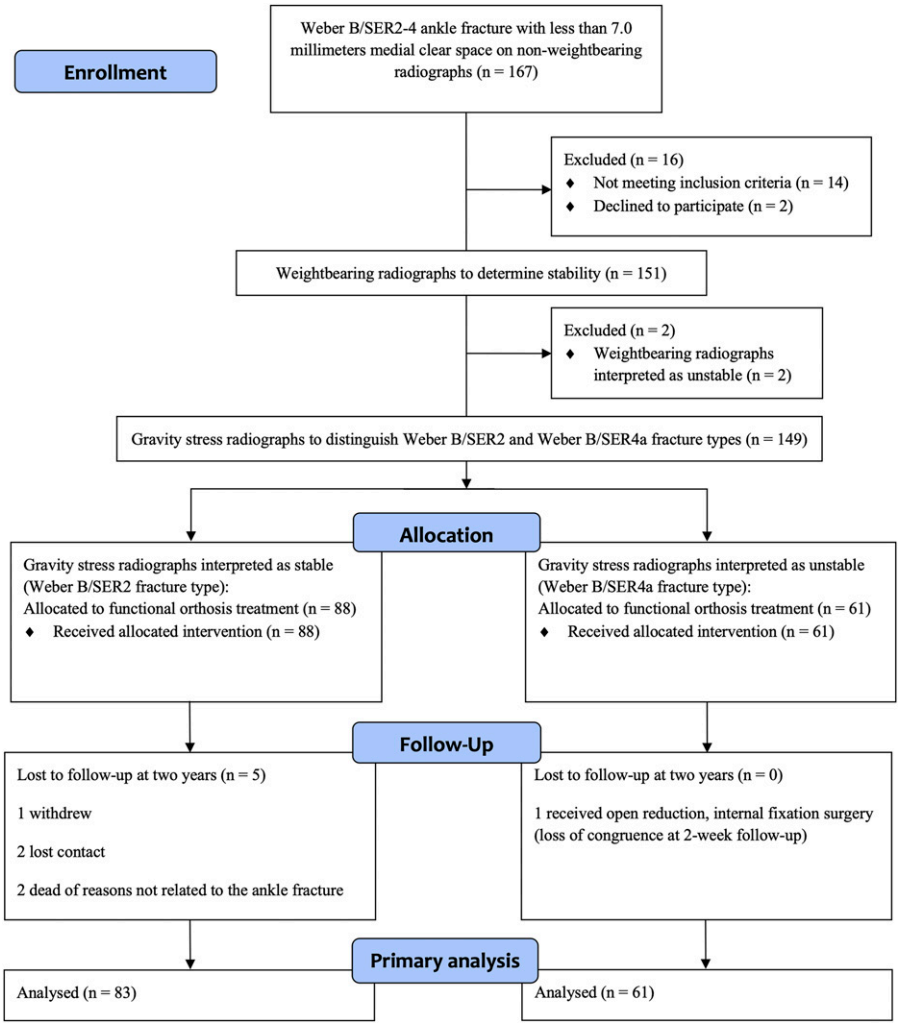

Fig. 1

Flow diagram for the study.

TABLE I Baseline Characteristics of 149 Patients Sustaining a Weber B Ankle Fracture with a Weight-Bearing Radiograph Considered Stable and Concurrent Gravity Stress Test Considered Stable (SER2) or Unstable (SER4a)

Characteristic	SER2 (N = 88)	SER4a (N = 61)
Age at injury* *(yr)*	53 ± 15	47 ± 15
Age ≥65 years†	25 (28%)	8 (13%)
Sex†		
Male	28 (32%)	32 (52%)
Female	60 (68%)	29 (48%)
Body mass index* *(kg/m²)*	26 ± 4	27 ± 5
Smoking status†		
Smoker	19 (22%)	15 (25%)
Nonsmoker	69 (78%)	46 (75%)

*The values are given as the mean and the standard deviation. †The values are given as the number of patients, with the percentage in parentheses.

MCS of the injured ankle at the 2-year follow-up compared with the uninjured side at baseline was −0.1 mm (95% CI, −0.2 to 0.1 mm; p = 0.392) for SER2 fractures and 0.0 mm (95% CI, −0.2 to 0.1 mm; p = 0.751) for SER4a fractures. The mean between-group difference comparing the SER2 and SER4a groups at the 2-year follow-up was −0.1 mm (95% CI, −0.3 to 0.1 mm; p = 0.311). These results were consistent with equivalence.

Safety

During the 2-year follow-up, 1 (0.7%) of 149 participants experienced a loss of tibiotalar congruence. This was detected at the 2-week follow-up in a patient in the SER4a group. Hence, the participant received standard operative treatment. Delayed fracture healing occurred in 1 (0.7%) of 149 participants, but the fracture healed without further intervention. This happened in the SER2 group. Furthermore, a symptomatic deep vein thrombosis occurred in 2 (1.3%) of 149 participants, 1 in each group, and pulmonary embolism occurred in 1 (0.7%) of 149 participants. The latter was in the SER2 group.

Discussion

In this study of Weber B/SER ankle fractures deemed to be stable on weight-bearing radiographs, treated with functional orthoses, and allowed to bear weight, those with a concurrent gravity stress test considered unstable (SER4a), indicating a partial deltoid ligament rupture, did not have unfavorable outcomes compared with those with a stable gravity stress test (SER2). Our results suggest that gravity stress tests are unnecessary if weight-bearing radiographs are stable.

Our results are broadly consistent with those in previous studies in that nonoperative treatment for Weber B fractures that are stable on weight-bearing (SER2 or SER4a) leads to satisfactory patient-reported and radiographic outcomes[2-4,10-13]. However, most studies have shown results for weight-bearing-stable fractures as a homogenous group and did not consider that weight-bearing-stable ankles may be entirely stable (SER2) or partially unstable (SER4a) as determined by stress tests. To our knowledge, a study by Seidel et al.[4] evaluating patient-reported outcomes and our study[3] on early radiographic results are the only 2 to make this distinction. Both found no apparent differences in outcomes between SER4a and SER2 fracture types. However, our previous study assessed radiographic results only, and the study by Seidel et al.[4] should be interpreted cautiously because of inconsistent treatment with different orthotic devices and immobilization.

TABLE II Patient-Reported Outcomes of 149 Participants with a Weight-Bearing-Stable Weber B Ankle Fracture*

	SER4a		Difference		SER2	
Ankle Function	Score*	No. of Patients	Mean†	P Value	Score*	No. of Patients
6 weeks						
MOXFQ	40 ± 18	59	−1.1 (−7.1 to 4.8)	0.717	41 ± 19	88
OMAS	55 ± 24	59	1.9 (−5.8 to 9.6)	0.622	53 ± 23	88
12 weeks						
MOXFQ	22 ± 16	58	−1.7 (−7.4 to 4.0)	0.560	24 ± 18	83
OMAS	73 ± 18	58	−0.6 (−7.0 to 5.9)	0.858	74 ± 21	83
1 year						
MOXFQ	7 ± 10	45	0.9 (−3.1 to 5.0)	0.648	6 ± 11	59
OMAS	93 ± 14	45	−0.8 (−5.8 to 4.1)	0.739	93 ± 10	59
2 years						
MOXFQ	3 ± 9	61	1.0 (−1.4 to 3.4)	0.397	2 ± 4	83
OMAS	97 ± 10	61	−0.3 (−3.2 to 2.6)	0.836	97 ± 6	83

*The values are given as the mean and the standard deviation. †The values are given as the mean, with the 95% CI in parentheses.

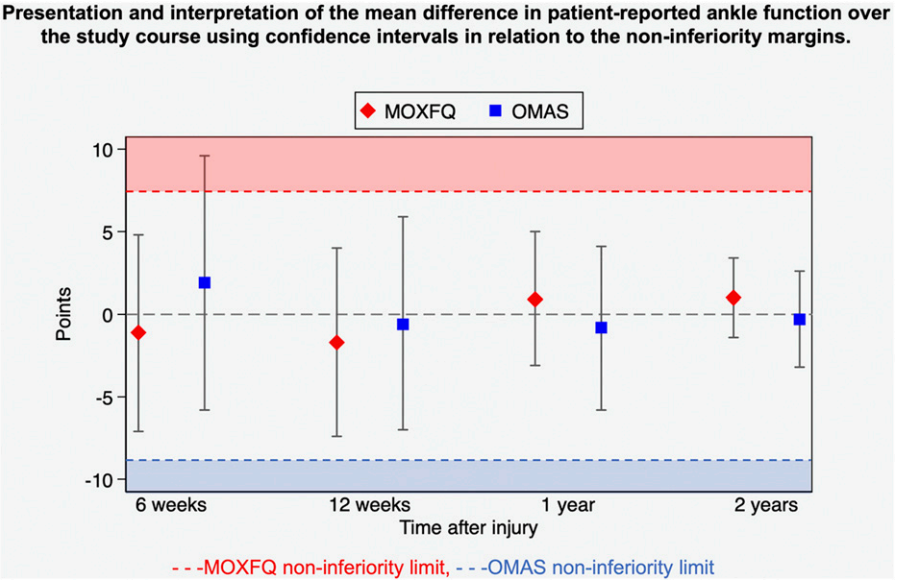

Fig. 2

Presentation and interpretation of the mean difference in patient-reported ankle function over the study course using 95% CIs in relation to the noninferiority margins.

Our study avoided some of the methodological shortcomings of previous studies that were primarily retrospective[10,12,13], lacking controls[10-13], or biased by inconsistent follow-up protocols[4], use of outcome measures lacking validity[10,11], or did not report a power calculation[4,10,11]. The present study prospectively enrolled a large sample of patients with comparable baseline characteristics and standard care for SER2 fractures as a comparator and had a high follow-up rate at 2 years. These factors enhance the generalizability of the results. In addition, the results were robust in sensitivity analyses.

Clinical Implications

We treated 3 (2%) of 151 patients operatively based on the results of weight-bearing radiographs. However, the number of operatively treated patients would have been 63 (42%) of 151 if gravity stress test results had been used as the indication for a surgical procedure. These numbers correspond with previous observations[2-4,10-13]. Based on the observed similar efficacy of the new nonoperative treatment for SER4a fractures compared with the same, established treatment for SER2 fractures in this study, the necessity of a surgical procedure for SER4a fractures seems questionable.

Moreover, advocates of nonoperative treatment for SER4a fractures have debated alternatives[2-4,6]. Some authors are apprehensive about functional options, arguing that movement and plantar flexion may disrupt deltoid ligament healing, causing potential chronic instability and elevated risk of posttraumatic arthritis. However, opposing data exist; Seidel et al.[4] demonstrated satisfactory clinical results at 2 years when using orthoses or boots. Comparably, we could not find unfavorable outcomes for SER4a compared with SER2 fractures treated functionally. Our findings of noninferior outcomes across the groups add to existing literature and support that all weight-bearing-stable fractures (SER2 or SER4a) may adhere to a consistent functional treatment protocol.

Limitations

Although no safety concerns were identified at the 2-year follow-up, we could not evaluate the potential risk of posttraumatic osteoarthritis at >2 years after fracture. Another limitation was the 7.0-mm MCS threshold that we used to determine instability on initial non-weight-bearing radiographs. This was shown to be useful in a previous study[4]. However, there is no evidence that ankles with an MCS widened to ≥7.0 mm on non-weight-bearing radiographs cannot still have an intact deep deltoid ligament and thus be stable. Therefore, we may have mistakenly excluded some partially unstable, SER4a fractures by interpreting them as unstable, SER4b fractures, which had the potential for detection bias based on unblinded radiographic assessment and could have influenced our results.

Conclusions

As replacement of gravity stress tests with weight-bearing radiographs as the surgical indication for Weber B/SER ankle fractures would lead to lower surgery rates, it is essential to recognize the effects of a concomitant unstable gravity stress test in weight-bearing-stable fractures treated nonoperatively. In this study, an unstable gravity stress test result (SER4a fracture) was not associated with worse patient-reported outcomes or more adverse events compared with a stable test (SER2) 2 years after functional orthosis treatment, suggesting that gravity stress tests are redundant if weight-bearing radiographs indicate stability.

Appendix

eA Supporting material provided by the authors is posted with the online version of this article as a data supplement at jbjs.org (http://links.lww.com/JBJS/H633). ■

NOTE: The authors acknowledge physiotherapists Åse-Helen Kristiansen and Silje K. Karlsen at the Department for Physical Medicine and Rehabilitation, Østfold Hospital Trust, who collected data and provided and cared for study participants. Furthermore, the authors thank Associate Professor Rene Holst, biostatistician at the Institute of Basic Medical Sciences, University of Oslo, for advice on statistical analyses and interpretation of data.

Martin G. Gregersen, PT, MSc[1,2]
Hilde Stendal Robinson, PT, PhD[2]
Marius Molund, MD, PhD[3]

[1]Department of Physical Medicine and Rehabilitation, Østfold Hospital Trust, Grålum, Norway

[2]Institute of Health and Society, Faculty of Medicine, University of Oslo, Oslo, Norway

[3]Department of Orthopaedic Surgery, Østfold Hospital Trust, Grålum, Norway

Email for corresponding author: martinggregersen@gmail.com

References

1. Meinberg EG, Agel J, Roberts CS, Karam MD, Kellam JF. Introduction: fracture and dislocation classification compendium-2018. J Orthop Trauma. 2018 Jan; 32(Suppl 1):S1-S170.

2. Dawe EJ, Shafafy R, Quayle J, Gougoulias N, Wee A, Sakellariou A. The effect of different methods of stability assessment on fixation rate and complications in supination external rotation (SER) 2/4 ankle fractures. Foot Ankle Surg. 2015 Jun; 21(2):86-90.

3. Gregersen MG, Molund M. Weightbearing radiographs reliably predict normal ankle congruence in Weber B/SER2 and 4a fractures: a prospective case-control study. Foot Ankle Int. 2021 Sep;42(9):1097-105.

4. Seidel A, Krause F, Weber M. Weightbearing vs gravity stress radiographs for stability evaluation of supination-external rotation fractures of the ankle. Foot Ankle Int. 2017 Jul;38(7):736-44.

5. Kortekangas T, Haapasalo H, Flinkkilä T, Ohtonen P, Nortunen S, Laine HJ, Järvinen TL, Pakarinen H. Three week versus six week immobilisation for stable Weber B type ankle fractures: randomised, multicentre, non-inferiority clinical trial. BMJ. 2019 Jan 23;364:k5432.

6. Gougoulias N, Sakellariou A. When is a simple fracture of the lateral malleolus not so simple? How to assess stability, which ones to fix and the role of the deltoid ligament. Bone Joint J. 2017 Jul;99-B(7):851-5.

7. McCormack DJ, Solan M, Aziz S, Faroug R, Kirmani S, Wright G, Mangwani J. Role of the posterior deep deltoid ligament in ankle fracture stability: a biomechanical cadaver study. World J Orthop. 2022 Nov 18;13(11):969-77.

8. Mittal R, Harris IA, Adie S, Naylor JM; CROSSBAT Study Group. Surgery for type B ankle fracture treatment: a Combined Randomised and Observational Study (CROSSBAT). BMJ Open. 2017 Mar 27;7(3):e013298.

9. Sanders DW, Tieszer C, Corbett B; Canadian Orthopedic Trauma Society. Operative versus nonoperative treatment of unstable lateral malleolar fractures: a randomized multicenter trial. J Orthop Trauma. 2012 Mar;26(3):129-34.

10. Weber M, Burmeister H, Flueckiger G, Krause FG. The use of weightbearing radiographs to assess the stability of supination-external rotation fractures of the ankle. Arch Orthop Trauma Surg. 2010 May;130(5):693-8.

11. Hoshino CM, Nomoto EK, Norheim EP, Harris TG. Correlation of weightbearing radiographs and stability of stress positive ankle fractures. Foot Ankle Int. 2012 Feb; 33(2):92-8.

12. Holmes JR, Acker WB 2nd, Murphy JM, McKinney A, Kadakia AR, Irwin TA. A novel algorithm for isolated Weber B ankle fractures: a retrospective review of 51 nonsurgically treated patients. J Am Acad Orthop Surg. 2016 Sep;24(9):645-52.

13. Rooney E, Ruse S, Talusan P, Holmes J, Walton D. Midterm follow-up of a novel algorithm for nonoperative Weber B ankle fractures. Foot Ankle Int. 2022 Nov; 43(11):1434-7.

14. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol. 2008 Apr;61(4):344-9.

15. Piaggio G, Elbourne DR, Pocock SJ, Evans SJ, Altman DG; CONSORT Group. Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. JAMA. 2012 Dec 26;308(24):2594-604.

16. Michelson JD, Varner KE, Checcone M. Diagnosing deltoid injury in ankle fractures: the gravity stress view. Clin Orthop Relat Res. 2001 Jun;(387):178-82.

17. Murphy JM, Kadakia AR, Irwin TA. Variability in radiographic medial clear space measurement of the normal weight-bearing ankle. Foot Ankle Int. 2012 Nov;33(11): 956-63.

18. Dawson J, Boller I, Doll H, Lavis G, Sharp R, Cooke P, Jenkinson C. The MOXFQ patient-reported questionnaire: assessment of data quality, reliability and validity in relation to foot and ankle surgery. Foot (Edinb). 2011 Jun;21(2):92-102.

19. Dawson J, Coffey J, Doll H, Lavis G, Cooke P, Herron M, Jenkinson C. A patient-based questionnaire to assess outcomes of foot surgery: validation in the context of surgery for hallux valgus. Qual Life Res. 2006 Sep;15(7):1211-22.

20. Jia Y, Huang H, Gagnier JJ. A systematic review of measurement properties of patient-reported outcome measures for use in patients with foot or ankle diseases. Qual Life Res. 2017 Aug;26(8):1969-2010.

21. Morley D, Jenkinson C, Doll H, Lavis G, Sharp R, Cooke P, Dawson J. The Manchester-Oxford Foot Questionnaire (MOXFQ): Development and validation of a summary index score. Bone Joint Res. 2013 Apr 3;2(4):66-9.

22. Dawson J, Boller I, Doll H, Lavis G, Sharp R, Cooke P, Jenkinson C. Minimally important change was estimated for the Manchester-Oxford Foot Questionnaire after foot/ankle surgery. J Clin Epidemiol. 2014 Jun;67(6):697-705.

23. Dawson J, Doll H, Coffey J, Jenkinson C; Oxford and Birmingham Foot and Ankle Clinical Research Group. Responsiveness and minimally important change for the Manchester-Oxford Foot Questionnaire (MOXFQ) compared with AOFAS and SF-36 assessments following surgery for hallux valgus. Osteoarthritis Cartilage. 2007 Aug; 15(8):918-31.

24. Arbab D, Kuhlmann K, Schnurr C, Lüring C, König D, Bouillon B. Comparison of the Manchester-Oxford Foot Questionnaire (MOXFQ) and the Self-Reported Foot and Ankle Outcome Score (SEFAS) in patients with foot or ankle surgery. Foot Ankle Surg. 2019 Jun;25(3):361-5.

25. Olerud C, Molander H. A scoring scale for symptom evaluation after ankle fracture. Arch Orthop Trauma Surg (1978). 1984;103(3):190-4.

26. Garratt AM, Naumann MG, Sigurdsen U, Utvåg SE, Stavem K. Evaluation of three patient reported outcome measures following operative fixation of closed ankle fractures. BMC Musculoskelet Disord. 2018 May 2;19(1):134.