The Intra- and Interrater Reliability of the Fragility Fractures of the Pelvis Classification and Their Therapeutic Relevance: A Multicenter Assessment

Philipp Pieroh, MD; Andreas Höch; Florian Gras, MD; Sven Märdian; Alexander Pflug; Silvan Wittenberg; Christoph Ihle; Notker Blankenburg; Kevin Dallacker-Losensky, MD; Tanja Lieselotte Schröder; Steven C. Herath, MD; Daniel Wagner, MD; Hans-George Palm, MD, MBA; Christoph Josten Prof; Fabian Stuby, MD
University Leipzig, Leipzig, Germany

Purpose: In the elderly, the fracture morphology and remaining stability of the pelvic ring differ. The fragility fractures of the pelvis (FFP) classification system was introduced to address these differences. However, data on the reliability and impact on treatment decision are still rare and were therefore determined in the present study.

Methods: 60 CT scans with FFP from 6 hospitals were classified (complete, only main group) 3 times by 6 experienced, 6 inexperienced surgeons, and 1 surgeon trained by the originator of the FFP classification ("gold standard"). Furthermore, the therapeutic decision (surgical/nonsurgical treatment) was determined. Intra-/interrater reliabilities were determined. The "gold standard", submitting hospital, and the majority vote were used as references to assess the percentage of agreement for each FFP main group was determined.

Results: For the FFP classification the intrarater reliability was moderate (complete: 0.46; main group: 0.60) and the interrater reliability was substantial for the main group (0.61) and moderate for the complete (0.53) classification. Percentage of agreement revealed values of 70% to 80%, depending on the reference used. The lowest agreement was found for FFP II and III. The intrarater (0.54) and interrater (0.42) reliability of the therapy decision were moderate. For FFP II, the lowest agreement of 66% was observed with respect to the therapy recommendation of the "gold standard". The observer´s therapy decisions were in agreement with those of the submitting hospital in >75% for FFP I only. For the remaining main groups lower agreements were observed, with the lowest value for FFP III (60%).

Conclusion: The FFP classification system, based on image morphological data only, seems to be a suitable indicator for the treatment decisions of FFP. However, the therapy of FFP II and III should be thoroughly evaluated. In addition, other clinical parameters seem to influence the therapy decision, independent of the classification, and may be responsible for the low agreement of the therapy decision of the submitting hospital with those of the observers.