ORIGINAL ARTICLE

Is there any purpose in classifying subtrochanteric fractures? The reproducibility of four classification systems

P. M. Guyver · M. J. H. McCarthy · Neil P. M. Jain · R. J. Poulter · C. J. P. McAllen · J. Keenan

Abstract

Introduction Classification systems are used for communication, planning treatment options, predicting outcomes and research purposes. The majority of subtrochanteric fractures are now treated with intramedullary nails and therefore questioning the need for classification.

Objectives To assess the intra- and inter-observer reproducibility of the Seinsheimer, AO and Russell-Taylor (RT) classification systems and to assess a new simple system (MCG).

Materials and methods The MCG system was developed to alert the surgeon to potential hazards: type 1—subtrochanteric fracture (ST#) with intact trochanters, type 2—ST# involving greater trochanter (entry point for nailing difficult), and type 3—ST# involving lesser trochanter (most unstable). Thirty-two anteroposterior and lateral radiographs of subtrochanteric fractures were classified independently for each of the 4 classification systems by 4 observers on 2 separate occasions.

Results The intra- and inter-observer variation was poor in all systems (highest Kappa 0.35). MCG had the best reproducibility followed by RT, then AO and Seinsheimer. The data were re-analysed to determine whether the findings were due to the presence of too many subgroups and whether the observers could more accurately identify important individual subclassifications: Seinsheimer 3a, AO31-A3.1, RT 1 or 2, RT a or b, and MCG3. The MCG3 had the narrowest ranges for intra- and inter-observer reproducibility.

Conclusions The classification systems analysed in this study have poor reproducibility and seem to be of little value in predicting the outcome of intramedullary nailing as all of the fractures achieved union. The MCG system may be of some use in alerting the surgeon to potential problems.

Keywords Classification systems · Reproducibility · Subtrochanteric fractures

P. M. Guyver · M. J. H. McCarthy · R. J. Poulter ·
C. J. P. McAllen · J. Keenan
Derriford Hospital NHS Trust Trauma and
Orthopaedic Department, Plymouth Hospitals,
Derriford Road, Plymouth, Devon PL6 8DH, UK

N. P. M. Jain (✉)
4 Parklands Drive, Askam-in-Furness, Cumbria LA16 7JP, UK
e-mail: npmjain@hotmail.com

Introduction

The perfect classification system needs to be reproducible [1] in order to aid accurate communication between surgeons and standardise research. In turn, good reproducibility will influence the overall reliability of that classification. In addition, the perfect classification system should guide the surgeon in his treatment options and predict outcome.

There have been numerous classification systems for subtrochanteric fractures introduced over time as treatment options have been modified (Seinsheimer (1978) [2]; Russell-Taylor classification (1984) [3]; and AO classification [4]). See Figs. 1, 2, 3.

Studies have already suggested poor reproducibility of the Seinsheimer [2] and AO [4] classification systems concluding that they are inaccurate for classifying subtrochanteric fractures.

One study investigating the reproducibility of the Seinsheimer classification showed that only 13 out of 50 fractures
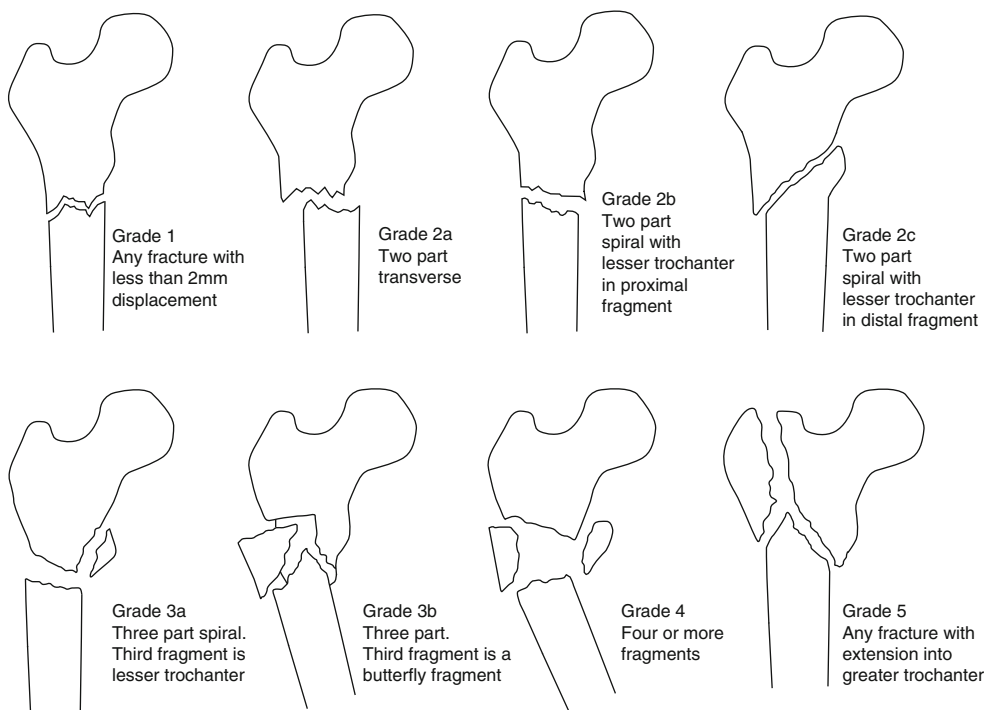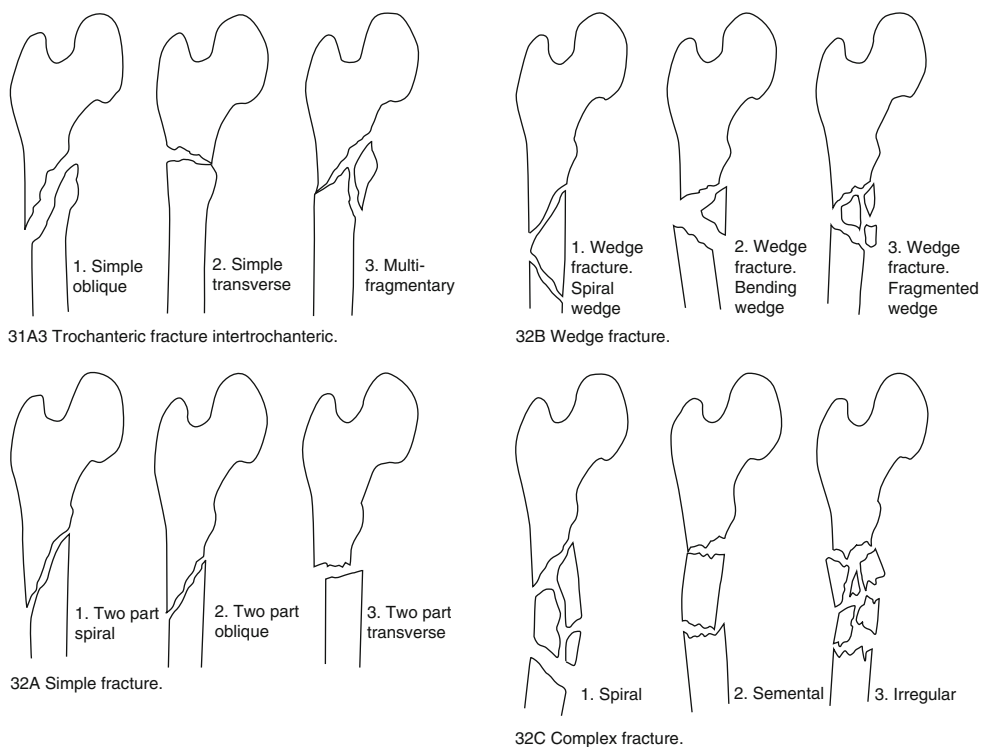
Fig. 1 Seinsheimer classification

Fig. 2 AO classification

(26%) were identically classified by all 4 observers, increasing to 62% when solely identifying the fractures classified as 3a [5]. This is important as the 3a subgroup was shown to have the highest rate of failure of fixation and persistent non-union in the original paper [2]. The complex AO classification system has been shown to be significantly more reliable when classifying by group as opposed to the more detailed subgroup [6].
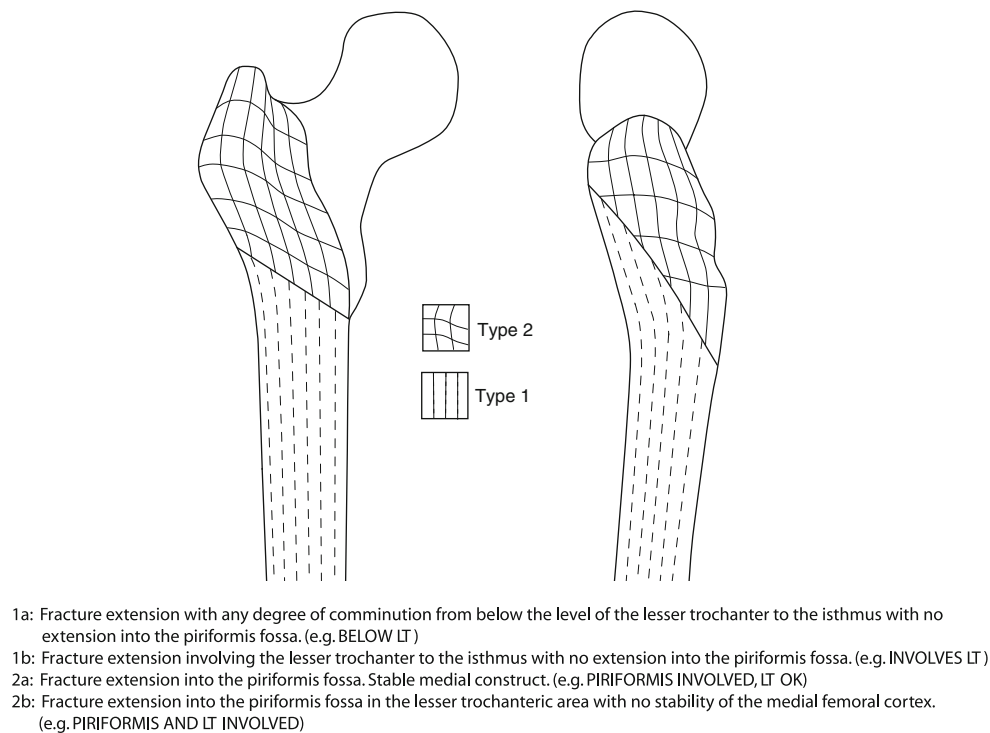
1a: Fracture extension with any degree of comminution from below the level of the lesser trochanter to the isthmus with no extension into the piriformis fossa. (e.g. BELOW LT)
1b: Fracture extension involving the lesser trochanter to the isthmus with no extension into the piriformis fossa. (e.g. INVOLVES LT)
2a: Fracture extension into the piriformis fossa. Stable medial construct. (e.g. PIRIFORMIS INVOLVED, LT OK)
2b: Fracture extension into the piriformis fossa in the lesser trochanteric area with no stability of the medial femoral cortex. (e.g. PIRIFORMIS AND LT INVOLVED)

Fig. 3 Russell-Taylor (RT) classification

A cohort of patients was followed up prospectively during the use of a new reconstruction nail—the Synthes© Proximal Femoral Nail Antirotation (PFNA). The poor reproducibility of subtrochanteric fractures classification systems was observed whilst classifying the fractures suitable for this nail use. We, therefore, developed a new classification system, 'MCG', as an attempt to improve the reproducibility by simplifying the classification of subtrochanteric fractures with the clinical aid of alerting the surgeon to potential hazards: type 1—subtrochanteric fracture with intact greater and lesser trochanters; type 2—subtrochanteric fractures involving greater trochanter, thus increasing the difficulty of obtaining the correct entry point for nailing; type 3—subtrochanteric fractures involving the lesser trochanter, the most unstable configuration and fractures in which reduction could be difficult to achieve. See Fig. 4.

The aim of this study was to assess the intra-observer and inter-observer reproducibility of the Seinsheimer [2], AO [4] and Russell-Taylor (RT) [3] classification systems along with the new system (MCG). Each classification system was noted to have varying numbers of subgroups Seinsheimer = 8, AO = 12, RT = 4 and MCG = 3. We did not use the classification systems during this study to guide treatment options as each fracture was treated with the same device (Synthes© PFNA).

Materials and methods

All patients treated for subtrochanteric fractures in our hospital between April 2006 and April 2007 were included in our study. Pathological fractures and those with inadequate quality radiographs were excluded leaving a total of 32 patients. All 32 anteroposterior (AP) and lateral radiographs were classified independently by 4 observers (2 Consultants and 2 Registrars) on two separate occasions. Each consultant had more than 10 years and each registrar more than 6 years experience in orthopaedics.

The radiographs were presented via a PowerPoint presentation (Microsoft Corp.) containing one patient's radiographs per slide (one AP, one Lateral). Each observer was provided with printed definitions and diagrams demonstrating the four classification systems of Seinsheimer, AO, RT and MCG. In the form of a test, the observers individually classified and recorded their selections for each fracture according to these classification systems. The observers were permitted to refer to their printed sheets throughout their classification of the fractures.

In order to prevent bias, the observers were not informed and there would be a second test. Six weeks later, the test was repeated but with the sequence of the slides altered. The observers were not informed that it was the same cohort of X-rays as they had observed previously.
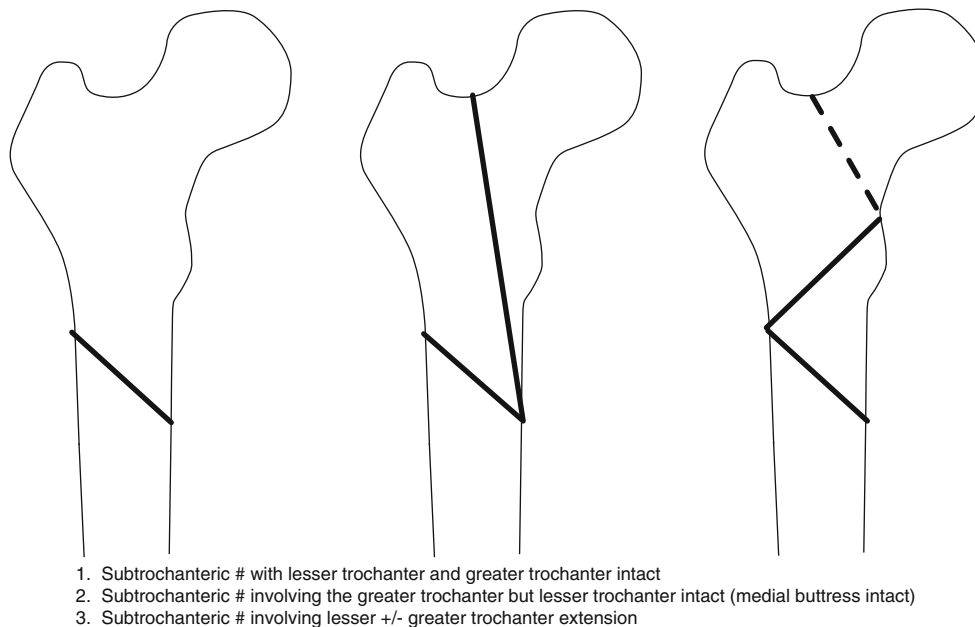
1. Subtrochanteric # with lesser trochanter and greater trochanter intact
2. Subtrochanteric # involving the greater trochanter but lesser trochanter intact (medial buttress intact)
3. Subtrochanteric # involving lesser +/- greater trochanter extension

Fig. 4 MCG classification

Kappa statistics were used to analyse the intra-observer and inter-observer reproducibility initially on each classification system.

In addition, Kappa statistics were also used to assess the reproducibility for certain individual classification grades of each system: Seinsheimer 3a, AO Group 3—subgroup 1 (31-A3.1), RT 1 or 2, RT a or b and MCG 3. The purpose of this element of the study was to simplify the classification systems and attempt to remove the inherent statistical bias that is observed when comparing classification systems that contain varying numbers of subgroups. The possibility is that the classification systems with fewer subgroups may only produce higher reproducibility results due to there being fewer options to choose from when reclassifying.

Kappa is a coefficient of agreement corrected for the probability of agreement by chance, which ranges from +1, representing perfect agreement, through 0, representing chance agreement, to −1, representing absolute disagreement. As defined by Landis and Koch, values of >0.80 are considered to be having almost perfect agreement; 0.61–0.80, substantial agreement; 0.41–0.60, moderate agreement; 0.21–0.40, fair agreement; 0–0.20, slight agreement; and 0, poor agreement [7]. SPSS® 11.0 software was used for the statistical analysis [8].

Results

We were unable to calculate Kappa values for both AO and Seinsheimer classification systems over the two sessions. This was due to the large number of subgroups and

Table 1 Assessment of all four classification systems reproducibility using mean kappa score

Mean kappa	Intra-observer	Inter-observer
Seins	NA very poor	NA very poor
AO	NA very poor	NA very poor
RT	0.25 fair	0.35 fair
MCG	0.30 fair	0.31 fair

relatively small sample size. For the RT classification, the mean kappa values were 0.25 for intra-observer variation and 0.35 for inter-observer variation. The MCG classification results showed a mean kappa value of 0.3 for intra-observer variation and 0.31 for inter-observer variation. See Table 1.

When analysing the individual classifications to see whether reproducibility improved, the MCG3 showed the best mean kappa value, 0.308, and the narrowest range (0.188–0.412) for intra-observer variation and 0.461 (range 0.446–0.629) for inter-observer variation. The RT 1 or 2 mean kappa value was 0.367 (range 0.271–0.675) for intra-observer variation and 0.387 (range 0.25–0.538) for inter-observer variation. RT a or b mean kappa was 0.296 (range 0.019–0.538) for intra-observer variation and 0.304 (range 0.043–0.592) for inter-observer variation. Seinsheimer 3a group showed a mean kappa value 0.295 (range −0.185–0.634) for intra-observer variation and 0.304 (range 0.004–0.796) for inter-observer variation. The AO31-A3.1 had the lowest mean kappa values, 0.075 (range 0.027–0.188) for intra-observer variation and 0.261 (range 0.125–0.428) for inter-observer variation. See Table 2.

Table 2 Assessment of reproducibility for certain individual classification grades of each system using mean kappa score

Mean kappa + range	Intra-observer	Inter-observer
SEINS 3A	0.295 (−0.185–0.634)	0.304 (−0.004–0.796)
AO 31-A3.1	0.075 (0.027–0.188)	0.261 (0.125–0.438)
RT 1 or 2	0.167 (0.271–0.675)	0.387 (0.25–0.538)
RT A or B	0.296 (0.019–0.538)	0.304 (0.043–0.592)
MCG3	0.308 (0.188–0.412)	0.461 (0.446–0.629)

All of these fractures were treated with the Synthes© Proximal Femoral Nail Antirotation (PFNA), and all but 2 achieved union. The 2 patients who had not united at 4 months were unfortunately lost to follow-up. We defined union based on clinical (no pain, tenderness or abnormal movement at the fracture site) and radiological (3 out of 4 bridging cortices) assessment [9]. Thus, the classification systems did not guide treatment or predict outcome or complications in this small study.

Discussion

The aim of this study was to assess the intra-observer and inter-observer reproducibility of the Seinsheimer [2], AO [4] and Russell-Taylor (RT) [3] classification systems along with a new system (MCG) specifically designed to improve upon the failings of the existing classification systems. Each system has its individual strengths and weaknesses.

The Seinsheimer system is widely used but has already been shown in the literature to have poor reproducibility. It is a descriptive system which offers many possible classification groups, although this may have contributed to its poor reproducibility.

The AO system is even more comprehensive with a greater number of subgroups, although again the benefit to its descriptive value potentially results in a poor reproducibility. Alternatively, the Russell-Taylor system has only 4 subgroups therefore making it less descriptive but possibly more reproducible. The new MCG system was designed specifically in order to be easy to use, be reproducible and be descriptive, primarily for clinical relevance rather than to purely describe the intricacies of fracture configuration. The objective being to create a classification system that would highlight to the surgeon the potential hazards involved in treating such fractures.

Our study agreed with the conclusions of those published by Gehrchen et al. [5] and Pervez et al. [6] who found improvements in the reproducibility of classification systems when observers only had to identify certain classification grades. However, at best, the mean agreement of all four classification systems was fair and when narrowed to certain classification grades, it only improved to moderate agreement. Interestingly, the observers still attained poor reproducibility despite having the classifications available to refer to throughout both tests.

Damany et al. [10] performed a Literature review of subtrochanteric classification publications from 1966 to 2003 and of 110 studies involving 2,725 fractures and 16 classification systems, and none were shown to be of value in determining treatment or for predicting outcome of subtrochanteric fractures. Specifically, we made a detailed comparison of existing classification systems with the addition of a new system that was designed with the explicit intention of addressing the criteria of a classification system. In spite of this, only a moderate inter-observer variation was demonstrated at best with the new classification (MCG).

We were able to test these classifications under standardised conditions with the use of digital radiographs. The small sample size resulted in statistical difficulty when analysing the classification systems with large numbers of subgroups (Seinsheimer and AO). However, given the uncommon nature of this injury and the relatively large size of our prospective cohort, this reflects poorly on the construct of those classification systems, i.e. systems used to predict outcome in uncommon conditions should be simple with minimal subgroups to permit adequate reproducibility.

All our fractures were treated with the same implant. Although this permitted a standardised assessment of outcome due to fracture type, it did not permit comparison of outcomes with different treatment modalities. However, we demonstrated a union rate of greater than 93% at 4 months. This may suggest that the vast majority of subtrochanteric femoral fractures can be treated with modern cephalomedullary nailing devices, thereby eliminating the need for the classification to guide treatment in these fractures.

In conclusion, the four subtrochanteric classification systems which we assessed were not found to be sufficiently reproducible to be of significant value in clinical practice.

References

1. Burstein AH (1993) Fracture classification systems: do they work and are they useful? J Bone Joint Surg Am 75:1743–1744
2. Seinsheimer F (1978) Subtrochanteric fractures of the femur. J Bone Joint Surg Am 60:300–306
3. Guyton JL (1998) Fractures of hip, acetabulum, and pelvis. In: Campbell's operative orthopaedics. 9th edn. St. Louis, Mosby, 2181–2276

4. Muller ME, Nazarian S, Koch P et al (1990) The comprehensive classification of fractures of long bones. Springer, Berlin
5. Gehrchen PM, Nielsen JO, Olesen B, Andresen BK (1997) Seinsheimer's classification of subtrochanteric fractures. Poor reproducibility of 4 observers' evaluation of 50 cases. Acta Orthop Scand 68(6):524–526
6. Pervez H, Parker MJ, Pryor GA et al (2002) Classification of trochanteric fracture of the proximal femur: a study of the reliability of current systems. Injury 33(8):713–715
7. Ramachandran M (2006) Basic orthopaedic sciences: the Stanmore guide. 1st edn. Hodder 13
8. SPSS® 11.0. SPSS Inc. (Nasdaq: SPSS) Chicago, IL, USA
9. Corrales LA, Morshed S, Bhandari M, Miclau T (2008) Variability in the assessment of fracture-healing in orthopaedic trauma studies. J Bone Joint Surg Am 90:1862–1868
10. Damany DS, Parker JM, Gurusamy K (2006) Classification of subtrochanteric fractures. Which method if any? J Bone Joint Surg Br 88-B:167