Does Computed Tomography Improve Reproducibility in the Classification of Transtrochanteric Fractures?*

A tomografia computadorizada melhora a reprodutibilidade na classificação das fraturas transtrocanterianas?

Murilo Alexandre¹ Giancarlo Cavalli Polesello¹,² Edio Cavassani Neto¹
Nayra Deise dos Anjos Rabelo¹ Marcelo Cavalheiro de Queiroz¹,² Walter Ricioli Junior¹,²

¹Hip Group, Orthopedics and Traumatology Department, Faculdade de Ciências Médicas da Santa Casa de São Paulo (FCMSCSP), Irmandade Santa Casa de Misericórdia, São Paulo, SP, Brazil
²School of Medical Sciences, Santa Casa de São Paulo, São Paulo, SP, Brazil

Address for correspondence Giancarlo Cavalli Polesello, MD, PhD, Rua Dr. Cesário Mota Junior, 112, Bairro Vila Buarque, São Paulo, SP, 01221-020, Brazil (e-mail: dot.quadril@hotmail.com).


**Abstract**

With the aging of the population, there was a significant increase in the prevalence of hip fractures, with high mortality rates, sequelae and expenses. Understanding the fracture profile and classifying it correctly is critical to define the appropriate treatment. Several radiographic classifications have been developed for transtrochanteric fractures, such as Tronzo, Evans-Jensen, AO and Boyd-Griffin, but their reproducibility is not always satisfactory. The present review aimed to elucidate whether the addition of computed tomography (CT) implies a greater reproducibility than simple radiography in the classification of transtrochanteric fractures, and whether this is a better examination to identify the fracture trait. A search was conducted in the PubMed, Lilacs, Scielo and Cochrane databases between July 2016 and June 2017, limited to the last 15 years. All retrospective, prospective and systematic reviews articles published in the English language, with evaluation of men and/or women, were considered for review. We have excluded case reports, studies that evaluated tomography or radiographs in isolation, and duplicate studies. The research presented 112 articles, of which 5 contemplated the proposed criteria. Reproducibility for the classification of transtrochanteric fractures presented variable results and was influenced by factors such as the type of classification, the use of the simplified or complete classification, the specialty of the evaluator, his experience, and the methodology proposed by the works. There are indications that there is benefit for the use of CT, especially for fractures considered unstable, but its use as a tool to ensure better reproducibility (intra-and interobserver) remains controversial and needs further studies.

**Introduction**

Transtrochanteric fractures are those between the base of the femoral neck up to 2.5 cm distal to the minor trochanter.¹ Although they may occur in young people after high-energy...
Computed tomography in transtrochanteric fractures classification

Alexandre et al.

Resumo

Com o envelhecimento populacional, houve um aumento significante da prevalência das fraturas do quadril, com alto índice de mortalidade, de sequelas, e alto custo. Compreender o perfil da fratura e classificá-la de forma correta é fundamental para definir o tratamento adequado. Diversas classificações radiográficas foram desenvolvidas para as fraturas transtrocanterianas, tais como as de Tronzo, de Evans-Jensen, de Boyd-Griffin e AO, porém sua reprodutibilidade nem sempre é satisfatória. O presente trabalho objetivou analisar se o acréscimo da tomografia computadorizada (TC) implica em maior reprodutibilidade do que a radiografia simples na classificação das fraturas transtrocanterianas e se esta é melhor para a identificação do traço de fratura. Foi realizada uma pesquisa nas bases de dados PubMed, Lilacs, Scielo e Cochrane entre julho de 2016 e junho de 2017, limitada aos últimos 15 anos. Todos os trabalhos retrospectivos, prospectivos e revisões sistemáticas publicados na língua inglesa, com avaliação de homens e/ou de mulheres, foram considerados para a revisão. Foram excluídos relatos de casos, estudos que avaliaram de forma isolada a TC ou radiografias e estudos duplicados. A pesquisa apresentou 112 artigos, dos quais 5 preenchiram os critérios propostos. A reprodutibilidade para a classificação das fraturas transtrocanterianas apresentou resultados variáveis e influenciados por fatores como o tipo de classificação, o uso da classificação simplificada ou completa, a especialidade do avaliador, a experiência e a metodologia proposta pelos trabalhos. Há indícios de que há algum benefício para o uso da TC, sobretudo para fraturas consideradas instáveis, porém sua utilização como ferramenta para garantir um melhor reprodutibilidade (intra- e interobservador) ainda permanece controversa e carece de mais estudos.

Palavras-chave

► fraturas do quadril
► classificação
► estudos de validação
► tomografia computadorizada
► por raios x
► radiografia

Trauma, elderly people falling to the ground are the most affected. Populational aging resulted in a significant increase in the prevalence of osteoporosis and of hip fractures, leading to higher mortality rates, sequelae, and associated costs when compared with the prevalence and the indexes of other fractures.

The diagnosis of transtrochanteric fractures is based on anamnesis and clinical evaluation; in an elderly individual, a low-energy trauma causes functional incapacity, shortening, and external rotation of the lower limb. Radiographic evaluation in anteroposterior (AP) (with lower limbs in internal rotation and traction) and lateral (L) views of the hip confirms the diagnosis, as well as the characteristics of the fracture.

A good understanding of the fracture type and its correct classification are critical to define the most appropriate osteosynthesis procedure. Thus, to be adequate for the daily clinical practice, the classification system must be simple, easy to apply, and present good intra- and interobserver reproducibility. In order to achieve this objective, several radiographic classifications have been developed, such as those by Boyd et al., Tronzo, by Jensen, by Evans, and the AO classification. However, in some cases, classificatory divergences impair intra- and interobserver reproducibility. These cases may require diagnostic methods for better identification and agreement to allow a good choice of treatment.

Computed tomography (CT) is a more advanced radiological technique that generates more detailed and complete images, which are useful for the evaluation of complex and articular fractures, such as those of the tibial pilon, of the tibial plateau, of the distal humeral, and of the calca-neus. Following this principle, the use of a tomographic classification was proposed for transtrochanteric fractures, as suggested by Nakano.

The availability of CT and its use led to the possibility of better intra- and interobserver reproducibility, as well as of a greater diagnostic precision for routine practice. However, the literature is still controversial. In some studies, this technique is deemed dispensable, whereas others considered it superior when compared with radiography.

Therefore, we have performed a systematic review to determine if the addition of CT presents superior intra- and/or interobserver reproducibility compared with fracture classifications based on plain radiographies alone.

Materials and Methods

A query was performed at the Pubmed, Lilacs, Scielo and Cochrane databases between July 2016 and June 2017, limited to the last 15 years. The following descriptors were used: (intertrochanteric OR hip) AND fracture AND (classification OR Tronzo OR AO OR Evans Jensen OR Boyd Griffin) AND (reproducibility OR validation studies OR reliability) AND (tomography OR CT) AND (x-ray OR radiography). At the initial query, all papers presenting two or more terms in the title and an abstract were included for review.

After the review of the papers, references were crossed to ensure the inclusion of all potential studies. All retrospective, prospective, and systematic English-language review articles with male and/or female patients comparing the reproducibility of radiographic and tomographic analysis of
transtrochanteric fracture classifications were considered for evaluation. We excluded case reports, studies assessing CT or radiography in isolation, duplicate studies, and diagnostic studies of occult or stress fractures.

The following data were extracted from the included studies: name of the first author, year of publication, country, study design, study objective, sample size, number of evaluators, as well as their specialties, classification used, method of analysis, and results. These data were tabulated in Microsoft Excel version 2016 (Microsoft Corporation, Redmond, WA, USA) and analyzed descriptively. The Preferred Reporting Items for Systematic Reviews and Meta-Analyzes (PRISMA) recommendation for systematic reviews was followed.34

The reproducibility of the studies was interpreted according to the kappa coefficient, based on the guidelines proposed by Landis et al: values from 0.00 to 0.20 indicate bad reproducibility; from 0.21 to 0.40, reasonable; from 0.41 to 0.60, moderate; from 0.61 to 0.80, strong; and from 0.81 to 1.00, excellent reproducibility.35

**Results**

Initially, the query resulted in 112 papers; after applying the inclusion and exclusion criteria, 5 papers were obtained. Published between 2003 and 2017, these papers present a minimum of 30 and a maximum of 110 patients, with a mean age ranging from 59 to 85 years old. Due to the restricted number of papers, conflicts of interests were not considered as exclusion criteria (►Fig. 1).

All of the papers used the AO classification system, while two of them employed the Evans-Jensen classification, and only one used the Boyd Griffin system (as evidenced in table 1).

Three papers concluded that the use of CT has benefits for the correct classification of transtrochanteric fractures with good reproducibility, especially in the identification of comminution and of unstable traits.29,30,33 In contrast, two papers indicated that CT does not present better classification reproducibility when compared with plain radiography1,33 (►Table 1).

**Discussion**

The reproducibility of the classification of transtrochanteric fractures had variable results and was influenced by several factors, such as classification system, sample size, use of the AO classification in its simplified or complete form, the specialty of the evaluators (radiologist versus orthopedist), the experience of the evaluators (staff professionals versus residents), and the methodology proposed by each work. Thus, the discussion was based on questions about the various factors that could influence these results.

---

**Fig. 1** Preferred Reporting Items for Systematic Reviews and Meta-Analyzes protocol for literature review.
Which classification has better reproducibility?
The simplified AO classification had better reproducibility compared with other classifications, such as the complete AO classification and its subgroups, the Evans-Jensen and Boyd Griffin classifications, both for tomography and radiography. These results are probably due to the lower number of information in the simplified AO classification, which would make both their comprehension and memorization easier when compared with the complete classification.

When analyzing CT scans, Ito et al. obtained a 75% concordance with the complete AO classification, and a 65.9% concordance with the Evans-Jensen classification. Van Embden D. et al. did not observe a higher reproducibility for the simplified AO classification when adding CT to the radiographic study, obtaining kappa values of 0.70 (strong) not using CT, and of 0.68 (strong) using CT images. Cavaignac E. et al. observed a reasonable reproducibility with the addition of CT to the complete AO classification (0.28 to 0.33), whereas the reproducibility of the Evans-Jensen classification was moderate (0.50) with only radiography, and reasonable (0.35) using radiography associated with CT.

Isida R. et al. found a strong reproducibility using CT and the complete AO classification (kappa value of 0.78), with a 100% positive predictive value, a 79% negative predictive value, and 95% specificity for comminution presence. When analyzed only by radiographs, the sensitivity was 48%, and the negative predictive value was 29%.

Does radiographic and tomographic standardization influence reproducibility?
Although CT has a well-documented importance in complex and intra-articular fracture detection, the results of transtrochanteric fractures may have been influenced by the divergences in the radiological techniques used.

Cavaignac E. et al. and Shen et al. analyzed AP and L hip radiographs in only 56% and 68% of the cases, respectively. This is an inadequate method, since Koval et al. previously demonstrated that the radiographic evaluation of transtrochanteric fractures should follow a pattern, obtaining two radiographic views (AP and L) in all cases. It is undeniable that radiographic examinations are limited in trauma rooms and in patients with pain and difficulty in positioning, but the creation of protocols for test standardization is of paramount importance.

The same occurs with CT, despite some standardization in its performance. A technique-related bias was observed by Cavaignac E et al., who used only axial CT sections, whereas Van Embden D. et al. used three tomographic views (coronal, sagittal and axial), but no 3D reconstruction. An additional factor is the lack of specification of the technique employed in the other studies, such as the number of radiographs, radiographic and CT scan views, patient positioning, and the use of 3D reconstruction models.

It can be inferred that the limitations of the tests and inadequate execution may have impaired the evaluation and generated a bias in the reproducibility results found.

Which classification subgroup is subject to greater variation? Does the use of computed tomography influence the choice of the treatment?
With the simplified AO classification, unstable fractures (A3) examined by CT presented higher reproducibility than those analyzed by plain radiographies.

The study by Van Embden D. et al.corroborates these results when CT is added to the analysis. In this study, type A3 fractures presented lower concordance and a greater number of surgical implant modifications when compared with implants chosen based only on plain radiographic evaluations (Figs. 2–3 to 4). It is worth mentioning that type A3 fractures present the highest complication rates, such as pseudoarthrosis and fixation failure in up to 32% of the cases.

Although type A3 fractures are subject to the largest changes in their classification, Van Embden D. et al. showed that types A1 and A2 fractures also had a change in the choice of implants after performing CT scans in 11 of their 30 cases. In addition, Shen et al. showed that CT use resulted in a better understanding of the fracture, which promoted a statistically significant reduction in surgical time for intra-medullary nailing placement, regardless of the classification.
Fig. 2 Anteroposterior and lateral views of the hip, AO classification 31 A1.

Fig. 3 Anteroposterior and lateral views of the hip, AO classification 31 A2.

Fig. 4 Anteroposterior and lateral views of the hip, AO classification 31 A3.
Therefore, adding CT scans to unstable fractures workup is plausible because it leads to a better understanding of the fracture trait, facilitates the proper choice of the implant, and is economically less costly compared with the expenses associated with increased surgical time and/or complications. However, the question is whether this examination would also not be beneficial for simple, stable fractures, due to the possibility of diagnostic inaccuracies and, consequently, to wrong choices of implants and to the occurrence of complications.

Does the specialty and/or experience of the evaluators influence reproducibility values?
One of the difficulties in evaluating these papers is the divergence in the type and level of the specialty of the evaluators. Two studies allowed the comparison between radiologists and orthopedists. Chapman CB. et al.\textsuperscript{32} demonstrated a strong reproducibility for radiologists (0.67) and a moderate reproducibility for orthopedists (0.57) in the complete AO and in the Evans-Jensen classifications. Mihir Thanvi et al.\textsuperscript{29} presented a lower reproducibility among radiologists compared with orthopedists, both for the simplified and complete OA classifications using radiographic, tomographic, and tomographic analysis with 3D reconstruction.

Furthermore, as exemplified by Isida R. et al.\textsuperscript{30} there is a bias in reproducibility evaluation that is related to the specialty of the evaluators. Orthopedists achieved moderate reproducibility (0.45), whereas radiologists presented excellent reproducibility (0.94), but the former evaluated exclusively radiographies, and the latter only analyzed CT scans. So it does not allow us to conclude that the specialty influences reproducibility.

Another variable that could influence reproducibility is the experience of the evaluators. In this sense, there are studies with resident physicians presenting classification errors and lower reproducibility, especially in unstable fractures,\textsuperscript{20,33} as well as studies in which evaluation by less experienced professionals did not compromise the quality of the classification compared with more experienced personnel.\textsuperscript{29}

Van Embden D. et al.\textsuperscript{33} supported this latter claim by demonstrating that residents achieved similar reproducibility rates as both trauma surgeons and radiologists. However, Cavaignac E. et al.,\textsuperscript{20} while agreeing that less experienced evaluators do not compromise the study, demonstrate excellent interobserver reproducibility (0.85) for senior evaluators and strong reproducibility (0.79) for junior evaluators when CT was added.

Limitations
There is a large variation in the number of evaluators in the review papers (between 2 and 11 individuals), and in their specialties (radiologists, orthopedists, and residents from both specialties), setting different degrees of experience and of technical knowledge.

Discrepancies in sample size were also an important factor (30 to 110 patients). This isolated factor could justify variations in kappa values, since the confidence interval for interobserver agreement depends directly on the sample size and on the number of evaluators, which are inversely proportional.\textsuperscript{37} Finally, the use of different methods and of difficult standardizations compromised the data analysis and the interpretation of the results.

Final Considerations
There is evidence that CT use is beneficial, especially for fractures considered unstable. Routine CT addition as a tool to ensure better reproducibility (both intra- and interobserver), as well as for decision-making in transtrochanteric fracture treatment, remains controversial. We believe that it is necessary to develop studies with a better level of evidence and equivalent methodologies to elucidate the benefits of CT as a tool for the classification of transtrochanteric fractures.

Conflicts of interests
The authors have no conflicts of interests to declare.

References
1 Müller ME, Nazarian S, Koch P. Classification AO des fractures. Berlin: Springer-Verlag; 1987
10 Shin YS, Chae JE, Kang TW, Han SB. Prospective randomized study comparing two cephalomedullary nails for elderly intertrochanteric fractures: Zimmer natural nail versus proximal femoral nail antitrotation II. Injury 2017;48(07):1550–1557