



Intracapsular neck of femur fractures secondary to civilian gunshot injuries: an inter- and intra-observer agreement study on classification and treatment using the AO/OTA classification

Sithombo Maqungo^{1,2} · Andrew Nicol^{1,3} · Maritz Laubscher¹ · Kaylin Williams¹ · Simon Graham^{1,4,5,6} · Michelle Henry^{1,7} · Ntambue Kauta¹ · Kirsty Berry¹

Received: 12 March 2024 / Accepted: 31 May 2024
© The Author(s) 2024

Abstract

Purpose Numerous classification systems have been developed for neck of femur fractures, but none have been tested for reliability in gunshot injuries. Our primary objective was to assess the inter-observer and intra-observer reliability of the AO/OTA classification system when applied to intracapsular neck of femur fractures secondary to low-velocity civilian gunshots wounds (GSWs). Our secondary objective was to test the reliability of the AO/OTA classification system in guiding surgeon treatment choices for these fractures.

Patients and methods Eighteen reviewers (six orthopaedic traumatologists, six general orthopaedic surgeons and six junior orthopaedic fellows) were given a set of 25 plain radiographs and CT scans of femur neck fractures secondary to GSW. For each clinical case, all reviewers selected a classification as well as treatment option from a list of given options. Inter-observer reliability was measured at the initial classification. The exercise was repeated 10–12 weeks later by the same 18 reviewers to test intra-observer reliability.

Results The Fleiss kappa values indicate only slight agreement amongst raters, across all experience levels, for both injury classification and treatment. Intra-observer agreement was fair across all experience levels for both injury classification and treatment.

Conclusion The AO/OTA classification showed only slight reliability in classification of gunshot fractures of the femur neck. With only fair reliability, it also failed to guide surgical treatment thus rendering its routine use in daily clinical practice of questionable value.

Keywords Gunshot · Neck of femur · Treatment options · Classification · Reliability

✉ Sithombo Maqungo
sithombo@msn.com

Andrew Nicol
andrew.nicol@uct.ac.za

Maritz Laubscher
maritzl@msn.com

Kaylin Williams
kaylinkwilliams@gmail.com

Simon Graham
simonmatthewgraham@doctors.org.uk

Michelle Henry
mhmish@gmail.com

Ntambue Kauta
jimmykauta@gmail.com

Kirsty Berry
berry.kirst@gmail.com

¹ Orthopaedic Trauma Service, Division of Orthopaedic Surgery, University of Cape Town, Groote Schuur Hospital, University of Cape Town, H49 Old Main Building, Observatory, Cape Town 7925, RSA

² Trauma Care & Injury Prevention, Division of Global Surgery, University of Cape Town, Cape Town, RSA

³ Trauma Unit, Division of General Surgery, University of Cape Town, Cape Town, RSA

⁴ Liverpool Orthopaedic and Trauma Service, Liverpool University Teaching Hospital Trust, Liverpool, UK

⁵ Nuffield Department of Orthopaedics, Rheumatology & Musculoskeletal Sciences, Oxford Trauma and Emergency Care, University of Oxford, Oxford, UK

⁶ Liverpool School of Tropical Medicine, Liverpool, UK

⁷ Centre for Higher Education Development, University of Cape Town, Cape Town, RSA

Introduction

Gunshot fractures of the hip joint are relatively rare injuries with notoriously poor outcomes [1, 2]. No reference standard exists for the classification and treatment of these devastating injuries. A number of classification systems have been used for intracapsular fractures of the femur neck, but none have found universal acceptance due to overall poor reliability.

The AO/OTA classification is at present the most comprehensive classification system used [3]. It considers level of the fracture and degree of displacement as well as the angle of the fracture lines. Several studies have however shown it to have poor reliability [4, 5]. The Garden classification and Pauwels' classification are also widely used, but they also have the shortcoming of poor reliability [6, 7].

Previous neck of femur (NOF) fracture reliability studies have been performed on closed fractures, frequently from low energy falls. No inter-observer and intra-observer reliability studies have been performed on classification and treatment for NOF fractures following penetrating injuries, including civilian gunshot injuries. The rarity and complexity of these injuries, together with the potential for poor outcomes and associated morbidity, necessitate a further quest for evidence-based medicine approach.

Aims

We therefore set out to:

- Assess the inter- and intra-observer agreement between surgeons in the classification of these injuries in a high-volume clinical setting.
- Analyse its accuracy in guiding the choice of treatment.
- Determine the effect of clinician experience on level of agreement.

Methods

This observational study was performed using a fixed panel of 18 observers who answered a set of questions regarding classification and treatment by analysing X-rays and CT scans of 25 cases with NOF fractures secondary to civilian gunshot injuries. A case example is shown in Fig. 1. The reviewers included orthopaedic trauma specialists ($n=6$) and general orthopaedic specialists ($n=6$) as well as orthopaedic fellows in training ($n=6$). They were from a total of eight different institutions. Cases were extracted from a single institution's orthopaedic trauma database between 2016 and 2021.



Fig. 1 Case example

Each reviewer received the AO/OTA fracture classification reference. This consists of nine subtypes in total, based on location of the fracture type (Fig. 2). All the reviewers were blinded to the treatment subsequently received by each patient. For each clinical case, they selected a classification as well as treatment option from a list of given options. There was no time limit imposed in order to allow for accurate assessment.

The interpretation was done over 2 rounds (Time 1 and Time 2), 10–12 weeks apart, without reference to their previous selections. For the second round, the cases were

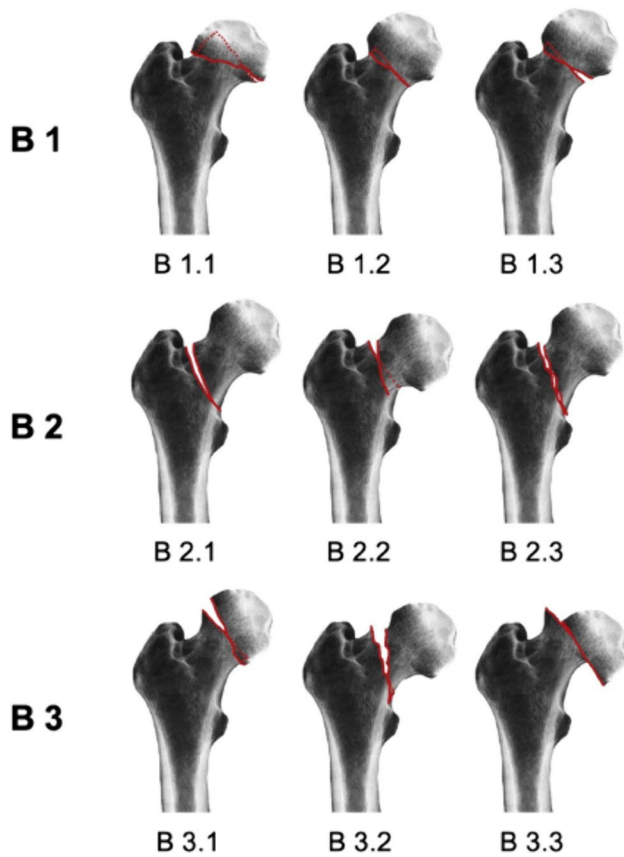


Fig. 2 AO/OTA classification

presented in a different order. The first-round classifications and treatment choices were used for inter-observer analysis and the second round for intra-observer analysis.

Study data were collected and managed using RED-Cap (Research Electronic Data Capture) electronic data capture tools.

Statistical analysis

Statistical analysis was performed by calculating the Cohen kappa value using SPSS 14.0 statistical software (IBM, Armonk, USA) for intra-observer reliability. In order to calculate the multirater kappa for inter-observer agreement, we used Fleiss kappa values.

We interpreted the kappa value coefficients according to the guidelines proposed by Landis and Koch: less than 0.00 equals poor reliability, 0.00 to 0.20 represents slight reliability, 0.21 to 0.40 fair reliability, 0.41 to 0.60 moderate reliability, 0.61 to 0.80 substantial agreement and 0.81 to 1.00 almost perfect agreement [8].

Table 1 Agreement before consolidation of AO OTA categories

Experience level	AO/OTA	Reliability	Treatment	Reliability
<i>Inter-observer agreement</i>				
All	0.087	Slight	0.031	Slight
Specialist trauma	0.067	Slight	0.042	Slight
General orthosurgeons	0.047	Slight	0.008	Slight
Fellows	0.110	Slight	0.003	Slight
<i>Intra-observer agreement</i>				
All	0.292	Fair	0.383	Fair
Specialist trauma	0.236	Fair	0.331	Fair
General orthosurgeons	0.378	Fair	0.464	Moderate
Fellows	0.262	Fair	0.380	Fair

Results

The Fleiss kappa values indicate only slight agreement amongst raters, across all experience levels, for both injury classification and treatment (Table 1). Intra-observer agreement was fair across all experience levels for both injury classification and treatment (Table 1).

For the total cohort, the inter-observer agreement for classification was 0.087 representing slight agreement. When broken down to the three subcategories based on experience, trauma surgeons had 0.067, general orthopaedic surgeons had 0.047 and fellows had 0.110 agreement, all representing slight reliability.

For the total cohort, the inter-observer agreement for treatment was 0.031 representing slight reliability. When broken down to the three subcategories, trauma surgeons had 0.042, general orthopaedic surgeons had 0.008 and fellows had 0.003 agreement, all representing slight reliability.

For the total cohort, the intra-observer agreement for classification was 0.292 representing fair reliability. When broken down to the three subcategories, trauma surgeons had 0.236, general orthopaedic surgeons had 0.378 and fellows had 0.262, all representing fair reliability.

For the total cohort, the intra-observer agreement for treatment was 0.383 representing fair reliability. When broken down to the three subcategories, trauma surgeons had 0.331 and fellows had 0.380, all representing fair reliability. With a rating of 0.464, only general orthopaedic surgeons demonstrated moderate reliability.

The most common classification types were B2.2 and B3.2 at both rounds of assessment (Time 1 and Time 2) (Fig. 3).

We then consolidated the fracture groups into B1, B2 and B3 without the subclassifications (Table 2). In this exercise, for the total cohort inter-observer agreement for classification, it was 0.146 representing slight reliability, signalling no

Fig. 3 Classification selections

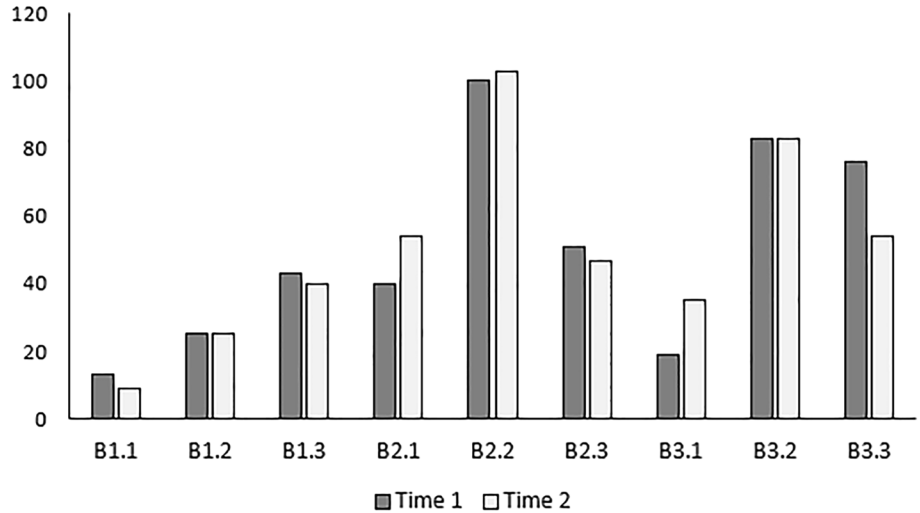


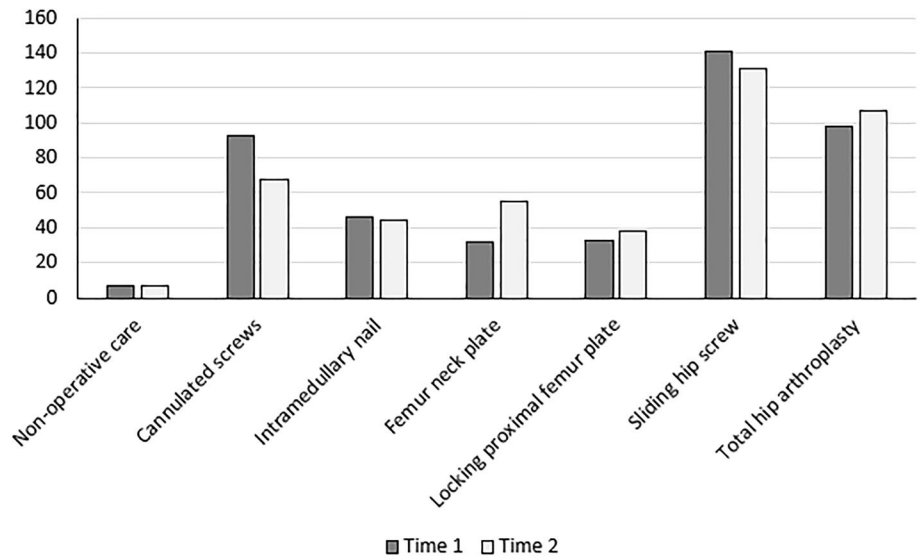
Table 2 Agreement after consolidation of AO OTA categories

Experience level	AO/OTA – 9 categories	Reliability	AO/OTA – 3 categories	Reliability
<i>Inter-observer agreement classification</i>				
All	0.087	Slight	0.146	Slight
Specialist trauma	0.067	Slight	0.130	Slight
General ortho-surgeons	0.047	Slight	0.130	Slight
Fellows	0.110	Slight	0.140	Slight
<i>Intra-observer agreement classification</i>				
All	0.292	Fair	0.436	Moderate
Specialist trauma	0.236	Fair	0.350	Fair
General ortho-surgeons	0.378	Fair	0.557	Moderate
Fellows	0.262	Fair	0.402	Fair

change when compared to the extended classification. Intra-observer agreement however improved slightly to 0.436 representing moderate reliability.

The three most common implant choices were sliding hip screw ($n=141$), total hip arthroplasty ($n=98$) and cannulated hip screws ($n=93$) at Time 1. At Time 2 observation, the top 3 remained the same but the order changed as follows: sliding hip screw ($N=131$), total hip arthroplasty ($n=107$) and cannulated screws ($n=68$). See Fig. 4.

Fig. 4 Treatment selections



Discussion

Gunshot fractures of the hip joint have notoriously poor outcomes, and when treated with internal fixation, they have high complication rates such as non-union, failure of fixation and avascular necrosis [9]. For hip fractures, the anatomical configuration and therefore classification generally determines the treatment option to be adopted. In this study, we assessed the commonly used AO/OTA classification for its inter- and intra-observer reliability in classifying gunshot fractures of the femur neck. We also assessed it for its reliability in guiding treatment choices. This is the first study to our knowledge to report on reliability of this classification in NOF fractures secondary to civilian gunshots. We have found only slight reliability amongst all experience levels when it comes to classification and fair reliability in guiding treatment options.

Ideally, a fracture classification system should have good inter-observer and intra-observer reliability and should also be able to provide information on stability, guide treatment interventions and allow for scientific comparisons of 'like with like'. It should also be able to predict anatomic and functional outcomes and be appropriate for daily clinical practice and audit [10, 11]. Femur neck fractures secondary to firearm injuries differ when compared to closed (commonly fragility) fractures due to the higher energy imparted and the inherent comminution that is present in all fractures.

Various classification systems have been proposed to classify intracapsular hip fractures, but none have found universal acceptance. The most commonly used system is that of Garden who divided them into four groups based on impaction or degree of displacement on anteroposterior radiographs [12]. Many subsequent studies however have doubted the value of the Garden system due to its poor reliability [4, 6, 13–18]. Parker was the first to show that the difference in the rates of fracture healing between Garden types III and IV was not sufficient to justify separating these two grades [14].

The Pauwel classification has also been used commonly. It has three subtypes, and it considers the angle of the fracture line relative to the femur shaft. It associated a greater vertical shear fracture line with an increase in incidence of non-union and malunion. It too however has been shown to have poor inter-observer reliability and has also been shown to be not predictive of non-union or avascular necrosis [7, 19]. Pauwel classification is also fraught with difficulties with accurate measuring of the fracture line angle due to rotation of the femur [20]. As these are penetrating injuries, often affecting younger patients compared to blunt trauma, applying the available classification systems has been challenging in the clinical setting.

The AO/OTA classification has also been found to not be reliable in both closed intracapsular and extracapsular fractures of the femur neck [21, 22]. In this study, we have

reached similar findings and a similar conclusion that it is too complicated for routine clinical use. Even when we collapse the subcategories and group together B1, B2 and B3 fractures without the subdivisions, the results remain the same, slight reliability, even though there was minor improvement, it was negligible to affect the rating. In previous studies, there has been an improvement in agreement rating when the AO classification was simplified into fewer categories [21]. This has not been the case in our study.

Reproducible and accurate fracture classification is important to guide the surgical implant of choice as well as the prognosis of the injury in terms of malunion, non-union and avascular necrosis. When one takes into account experience levels amongst the observers, only general orthopaedic surgeons could reach fair agreement on treatment, with many opting for a sliding hip screw device (Fig. 4). Prior to our current study, no agreement studies have been performed on treatment choices for these injuries. And it is clear from this data that the low reliability meant treatment choices were also unreliable as many surgeons changed their opinion of treatment choice during the second round.

The high proportion of total hip arthroplasty as a treatment choice was unexpected given the average age of 28 years for the cohort. There is no strong evidence to support this practice. Only sporadic case reports have reported on arthroplasty being performed much later in a staged manner, rather than in the acute setting [23–25].

Limitations

The low numbers are a recognised limitation of our study, but these are relatively rare injuries collected over an extended period. Our unit is a high-volume Level 1 Trauma Centre in an urban area with a high burden of gunshot injuries. All observers practised in the same country, albeit at different institutions, so the results may not be generalisable to other countries or regions.

Conclusion

We have found the AO/OTA classification to have only slight intra- and inter-observer reliability in classifying intracapsular civilian gunshot fractures of the femoral neck. The experience level of the reviewers did not improve its reliability. With only fair reliability, it also failed to guide surgical treatment thus rendering its routine use in daily clinical practice of questionable value.

Future research needs to focus on developing a reliable classification system for these injuries that is able to both guide treatment and to predict the outcome. [17].

Acknowledgements The authors would like to thank the following reviewers for their participation: Michael Abramson, Delroy Arnolds, Tsepo Bam, Craig Blake, Craig Brown, Kudzai Chironga, Ayik Goud Deng, Gian Du Preez, Danie Hugo, Fred Louw, Thamsanqa Mazibuko, Jeannie McCaul, Stewart Mears, Thivani Naidoo, Joseph Seritsane and Livan Menes Turino.

Funding Open access funding provided by University of Cape Town. This research has been funded in part by the National Research Foundation of South Africa, Grant No. 138208.

Declarations

Conflict of interest The authors certify that they have no affiliations with or involvement in any organisation or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership or other equity interest and expert testimony or patent-licensing arrangements) or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

Ethical approval Ethical approval for the study was granted by the University of Cape Town Human Research Ethics Committee. Approval number: 803/2021.

Human and animal participants The study does not involve the use of animals by any of the authors.

Consent for publication This is a radiological study of clinical images. At the time of treatment, all patients gave informed consent to use of clinical data and radiological images for research purposes.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Israel H, Cannada LK (2020) Gunshot wounds to the hip: doomed to failure? *J Surg Ortho Adv*. <https://doi.org/10.3113/JSOA.2020.0135>
2. Maqungo S, Fegredo D, Brkljac M, Laubscher M (2020) Gunshot wounds to the hip. *J Orthop* 22:530–534. <https://doi.org/10.1016/j.jor.2020.09.018>
3. Meinberg EG, Agel J, Roberts CS, Karam MD, Kellam JF (2018) Fracture and dislocation classification compendium-2018. *J Orthop Trauma* 32:S1–S170. <https://doi.org/10.1097/BOT.0000000000001063>
4. Masionis P et al (2019) The reliability of a garden, AO and simple II stage classifications for intracapsular hip fractures. *Orthop Traumatol Surg Res* 105(1):29–33. <https://doi.org/10.1016/j.otsr.2018.11.007>
5. van Embden D, Rhemrev SJ, Meylaerts SAG, Roukema GR (2010) The comparison of two classifications for trochanteric femur fractures: the AO/ASIF classification and the jensen classification. *Injury* 41(4):377–381. <https://doi.org/10.1016/j.injury.2009.10.007>
6. Van Embden D, Rhemrev SJ, Genelin F, Meylaerts SAG, Roukema GR (2012) The reliability of a simplified garden classification for intracapsular hip fractures. *Orthop Traumatol Surg Res* 98(4):405–408. <https://doi.org/10.1016/j.otsr.2012.02.003>
7. Van Embden D, Roukema GR, Rhemrev SJ, Genelin F, Meylaerts SAG (2011) The Pauwels classification for intracapsular hip fractures: Is it reliable? *Injury* 42(11):1238–1240. <https://doi.org/10.1016/j.injury.2010.11.053>
8. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* 33(1):159–174
9. Zhang Y et al (2020) Gunshot wounds to the hip: doomed to failure? *J Surg Ortho Adv* 29(3):135–140. <https://doi.org/10.3113/JSOA.2020.0135>
10. Pervez H, Parker MJ, Pryor GA, Lutchman L, Chirodian N (2002) Classification of trochanteric fracture of the proximal femur: a study of the reliability of current systems. *Int J Care Injured* 33:713–715
11. Audigé L, Bhandari M, Hanson B, Kellam J (2005) A concept for the validation of fracture classifications. *J Orthop Trauma* 19(6):404–409
12. Garden RS (1961) Low-angle fixation in fractures of the femoral neck. *J Bone Joint Surg* 43-B(4):647–663
13. Frandsen PA, Andersen E, Madsen F, Skjold T (1988) Garden's classification of femoral neck fractures. an assessment of inter-observer variation. *J Bone Joint Surg* 70-B(4):588–590
14. Parker M (1993) Garden grading of intracapsular fractures: meaningful or misleading? *Injury* 24(4):241–242
15. Kazley JM, Banerjee S, Abousayed MM, Rosenbaum AJ (2018) Classifications in brief: garden classification of femoral neck fractures. *Clin Orthop Rel Res* 476(2):441–445. <https://doi.org/10.1007/s11999.0000000000000066>
16. Zlowodzki M, Bhandari M, Keel M, Hanson BP, Schemitsch E (2005) Perception of garden's classification for femoral neck fractures: an international survey of 298 orthopaedic trauma surgeons. *Arch Orthop Trauma Surg* 125(7):503–505. <https://doi.org/10.1007/s00402-005-0022-4>
17. Beimers L et al (2002) Subcapital hip fractures: the garden classification should be replaced, not collapsed. *Can J Surg* 45(6):411–414
18. Thomsen N et al (1996) Observer variation in the radiographic classification of fractures of the neck of the femur using garden's system. *Int Orthop* 20:326–329. <https://doi.org/10.1007/s002640050087>
19. Parker MJ, Dynan Y (1998) Is pauwels classification still valid? *Injury* 29(7):521–523
20. Bartončiček J (2001) Special interest paper pauwels' classification of femoral neck fractures: correct interpretation of the original. *J Orthop Trauma* 15(5):358–360
21. Blundell CM, Parker MJ, Pryor GA, Hopkinson-Woolley J, Bhonsle SS (1998) Assessment of the AO classification of intracapsular fractures of the proximal femur. *J Bone Joint Surg* 80(4):679–683
22. Chan G et al (2021) Inter- and intra-observer reliability of the new AO/OTA classification of proximal femur fractures. *Injury* 52(6):1434–1437. <https://doi.org/10.1016/j.injury.2020.10.067>
23. Zandi R, Talebi S, Ehsani A, Nodehi S (2023) Two-step sequential management for hip arthroplasty after hip joint gunshot injury: a case report. *Clin Case Rep*. <https://doi.org/10.1002/ccr3.7569>
24. Naziri Q et al (2013) Posttraumatic arthritis from gunshot injuries to the hip requiring a primary THA. *Orthopedics* 36(12):e1549–e1554
25. Bell C, Skibicki HE, Post ZD, Ong AC, Ponzio DY (2022) Gunshot wound resulting in femoral neck fracture treated with staged total hip arthroplasty. *Arthroplast Today* 14:44–47. <https://doi.org/10.1016/j.artd.2021.12.010>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.