


RESEARCH ARTICLE

Open Access



# Development of a radiographic scoring system for new bone formation in gout

Chang-Nam Son<sup>1,2\*</sup> , Ken Cai<sup>2</sup>, Sarah Stewart<sup>2</sup>, John Ferrier<sup>3</sup>, Karen Billington<sup>3</sup>, Yun-Jung Jack Tsai<sup>3</sup>, Thomas Bardin<sup>4</sup>, Anne Horne<sup>2</sup>, Lisa K. Stamp<sup>5</sup>, Anthony Doyle<sup>6</sup> and Nicola Dalbeth<sup>2</sup>

## Abstract

**Background:** Features of new bone formation (NBF) are common in tophaceous gout. The aim of this project was to develop a plain radiographic scoring system for NBF in gout.

**Methods:** Informed by a literature review, scoring systems were tested in 80 individual 1st and 5th metatarsophalangeal joints. Plain radiography scores were compared with computed tomography (CT) measurements of the same joints. The best-performing scoring system was then tested in paired sets of hand and foot radiographs obtained over 2 years from an additional 25 patients. Inter-reader reproducibility was assessed using intraclass correlation coefficients (ICC). NBF scores were correlated with plain radiographic erosion scores (using the gout-modified Sharp-van der Heijde system).

**Results:** Following a series of structured reviews of plain radiographs and scoring exercises, a semi-quantitative scoring system for sclerosis and spur was developed. In the individual joint analysis, the inter-observer ICC (95% CI) was 0.84 (0.76–0.89) for sclerosis and 0.81 (0.72–0.87) for spur. Plain radiographic sclerosis and spur scores correlated with CT measurements ( $r = 0.65$ – $0.74$ ,  $P < 0.001$  for all analyses). For the hand and foot radiograph sets, the inter-observer ICC (95% CI) was 0.94 (0.90–0.98) for sclerosis score and 0.76 (0.65–0.84) for spur score. Sclerosis and spur scores correlated highly with plain radiographic erosion scores ( $r = 0.87$  and  $0.71$  respectively), but not with change in erosion scores over 2 years ( $r = -0.04$ – $0.15$ ).

**Conclusion:** A semi-quantitative plain radiographic scoring method for the assessment of NBF in gout is feasible, valid, and reproducible. This method may facilitate consistent measurement of NBF in gout.

**Keywords:** Gout, New bone formation, Radiography

## Introduction

Features of new bone formation such as sclerosis and spurs are common on plain radiography in tophaceous gout [1]. While a plain radiographic damage scoring system incorporating bone erosion and joint space narrowing has been developed [2], there are no published methods for scoring new bone formation in gout.

Although there have been some reports of reparative processes with increased new bone formation in response to intensive urate-lowering therapy [3, 4], it is not possible to systematically quantify these changes at present, due to the lack of a validated scoring method. The aim of this project was to develop a plain radiographic scoring system for new bone formation in patients with gout.

## Materials and methods

Plain radiographs and computed tomography (CT) scans were obtained for research, approved by the New Zealand Ministry of Health (NTX/08/06/050/AM06 and MEC/11/06/060). The patients were diagnosed on

\*Correspondence: cnson@kmu.ac.kr

<sup>1</sup> Keimyung University School of Medicine, Daegu, South Korea  
Full list of author information is available at the end of the article



© The Author(s) 2021. **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/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

the basis of the 2015 American College of Rheumatology/European League Against Rheumatism gout classification criteria [5]. All participants provided written informed consent. Previously published definitions of new bone formation features in gout were used in this project [1]: ankylosis—fusion of the bones of a joint, with trabeculae crossing the joint space; osteophyte—bone projection arising along the joint margin and associated with cartilage; periosteal new bone formation—bone proliferation arising from the periosteum; sclerosis—increased density of medullary or subcortical bone; and spur—a sharp spicule of dense bone proliferation extending at an acute angle from the cortex.

### Literature review of scoring systems

We initially performed a scoping literature review of scoring systems for new bone formation in bone and joint diseases, searching in MEDLINE (via PubMed) and EMBASE databases, published from January 1, 1946, to February 28, 2020. The first search was performed on April 29, 2019, and repeated on February 28, 2020, to capture any additional studies published since the original search was performed. Details of each unique scoring system and the numbers of papers reporting each scoring system were extracted (Supplementary Table 1). The protocol was registered on PROSPERO (CRD: 42019138077; 28-08-2019).

### Initial testing of new bone formation scoring system in individual joints

Two rheumatologists (KC and ND) and a musculoskeletal (MSK) radiologist (AD) undertook a structured review of plain radiographs from patients with gout. Informed by the literature review, a range of scoring systems including semi-quantitative scoring systems used in osteoarthritis (Oslo Hand Osteoarthritis MRI Score [6] and Boston Leeds Osteoarthritis Knee Score [7]) and psoriatic arthritis (proliferation score of the Psoriatic Arthritis Ratingen Score (PARS) [8]) were chosen for scoring (Supplementary Table 2). In addition, quantitative measurement of lesions in millimeters and the ratio of the length of the largest lesion relative to the width of the adjacent joint surface in millimeters was selected. In an inter-reader reproducibility exercise, a rheumatologist (KC) and MSK radiologist (JF) scored 80 individual joints (40 1st metatarsophalangeal (MTP) joints and 40 5th MTP joints) from 20 separate patients with gout using these methods (see Supplementary Table 3 for clinical features of these participants).

### Refinement and retesting of new bone formation scoring system in individual joints

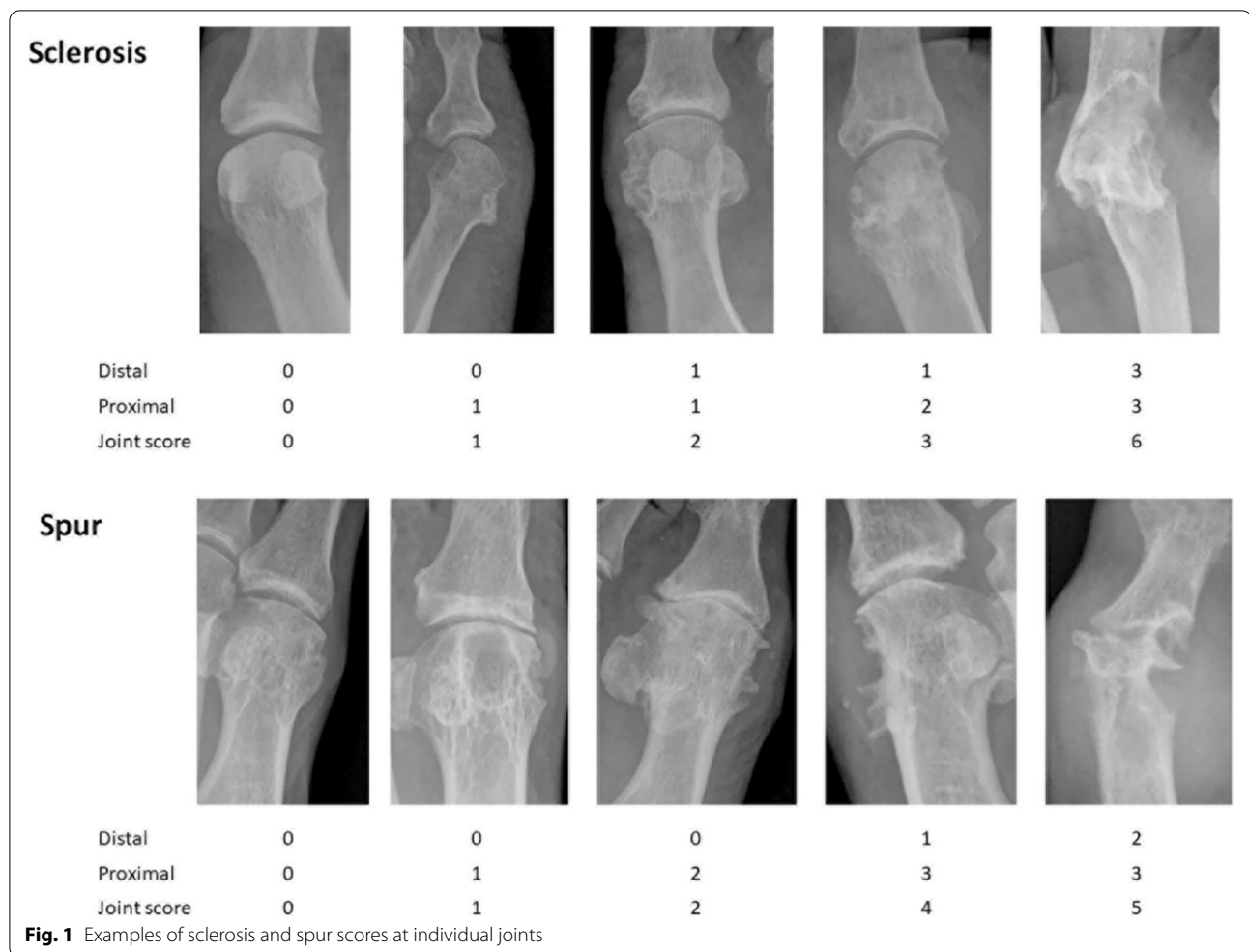
Following these initial results, we proceeded to a second structured review of a new set of plain radiographs (AD, KC, and ND). This work included standardization of definitions, approaches, and projections and development of reference images for scoring (Fig. 1). In a further inter-reader reproducibility exercise, two rheumatologists (CS and ND) rescored the same 80 individual joints (40 1st MTP joints and 40 5th MTP joints) for sclerosis and spur using the new scoring system, modified from the Kannus semi-quantitative plain radiographic scoring system for sclerosis in knee osteoarthritis [9] and the Thumb Base Osteoarthritis Magnetic Resonance Imaging Scoring System (TOMS) respectively [10].

### Comparison with CT measurements of individual joints

Construct validity was assessed by correlating the plain radiography scores with CT measurements of sclerosis and spur at the same individual joints, as previously described [1]. CT is considered the reference standard for the assessment of bone structural changes in arthritis [11, 12] and allows measurement in three dimensions. A MSK radiologist (YJT), who was blinded to all plain radiographic scoring, analyzed the CT scans obtained on the same date as the plain radiographs. The following sclerosis data were measured at the same 80 individual joints (40 1st MTP joints and 40 5th MTP joints): the presence of sclerosis, Hounsfield units at the site of maximum sclerosis, and the ratio of the depth of sclerosis relative to the width of the adjacent joint surface (in any dimension). The following spur data were measured at the same joints: the presence of spur, length of the largest spur in millimeters (in any dimension), and the ratio of the depth of sclerosis relative to the width of the adjacent joint surface (in any dimension).

### Testing of scoring system in hand and foot radiographs

The scoring system was then tested in paired sets of hand and foot radiographs over a 2-year period from a further 25 patients with gout receiving treatment to serum urate target ( $< 0.36$  mmol/L) gout management (at baseline, mean (standard deviation (SD)) disease duration 22 (11) years, 72% with subcutaneous tophi, mean (SD) allopurinol dose at baseline 296 (149) mg/day, with dose titration to 430 (185) mg/day at year 2, serum urate 0.43 (0.11) mmol/L at baseline, reducing to 0.32 (0.06) mmol/L at year 2; see Supplementary Table 4 for detailed clinical characteristics of these participants). In an inter-reader reproducibility exercise, the hand and foot radiographs were scored by two rheumatologists (CS and ND), using the same regions scored in the gout-modified Sharp-van



der Heijde scoring system for bone erosion [2, 13] (hand distal interphalangeal joints, proximal interphalangeal joints, interphalangeal (IP) joint of the thumb, metacarpophalangeal joints, base of the first metacarpal, multangular, scaphoid, lunate, radius, ulna, MTP joints, and IP joints of the hallux). The proximal and distal bones were scored separately for joints with long bones (maximum score 6 for each joint), and single bones for the carpal and tarsal bones (maximum score 3 for each bone). The same paired sets of hand and foot radiographs were scored for the gout-modified Sharp-van der Heijde erosion scores [2, 13] by a MSK radiologist (KB). All readers were blinded to each other's scores. Paired sets of radiographs were scored in a known order.

**Statistical analysis**

Inter-reader reliability was assessed by intraclass correlation coefficients (ICCs) with 95% confidence intervals (95% CI). Median and interquartile range (IQR) was used to describe new bone formation scores. Correlations

between the new bone formation scores (average of two scorers) and other modalities/scoring systems as well as clinical characteristics were assessed using Spearman's correlation coefficients. Differences between the groups were analyzed using the Mann-Whitney *U* and Wilcoxon signed-rank tests. All statistical analyses were performed using SPSS (version 25.0, IBM Corp., Armonk, NY, USA).

**Results**

**Literature review of scoring systems**

There were 840 papers identified, reporting 86 unique scoring systems for new bone formation (Supplementary Fig. S1). These scoring systems were applicable to a range of musculoskeletal conditions including osteoarthritis (*n* = 61 scoring systems), spondyloarthropathies (*n* = 16), diffuse idiopathic skeletal hyperostosis (*n* = 2), heterotopic ossification (*n* = 2), rheumatoid arthritis (*n* = 2), psoriatic arthritis (*n* = 1), slipped capital femoral epiphysis (*n* = 1), and juvenile idiopathic arthritis (*n* = 1). New bone formation was scored either as part of a composite

score ( $n = 40$  unique scoring systems) or as a stand-alone score ( $n = 44$ ) or a combination of both composite and stand-alone scoring systems ( $n = 2$ ). Those involving assessment of the individual features of new bone formation used semi-quantitative scales (i.e., none, mild, medium, severe) ( $n = 41$ , 89%). No scoring system used a quantitative scale. The most commonly used scoring systems are shown in Supplementary Table 1.

#### Initial testing of new bone formation scoring system in individual joints

Following the first structured review of plain radiographs, 80 MTP joints were scored by two readers based on osteoarthritis and psoriatic scoring systems, as well as quantitative measurement of lesions (Supplementary Table 2). Inter-reader ICCs were low (0.23–0.64 for osteophyte, 0.35–0.54 for sclerosis, and  $-0.05$  to 0.39 for spur) (Supplementary Table 5). No periosteal bone formation was scored by either reader, and one joint was scored for ankylosis by both readers.

#### Refinement and retesting of the new bone formation scoring system in individual joints

In the second structured review of plain radiographs, definitions, approaches, and projections for scoring were discussed and standardized. As sclerosis and spurs were the most common lesions specific to gout, we focused on developing a scoring system for these features, noting that although osteophytes are common in joints affected by gout, they are non-specific. Additionally, a scoring system for periosteal new bone formation and ankylosis was not progressed as these lesions were so uncommon. The low inter-reader reproducibility using the PARS system also led us to develop separate scores for sclerosis and spur, rather than a single new bone formation score.

The Kanner semi-quantitative plain radiographic scoring system for sclerosis in knee osteoarthritis [9] was modified for use in gout, as follows: 0 = no sclerosis, 1 =  $< 1/3$  bone affected, 2 =  $1/3$ – $2/3$  bone affected, and 3 =  $> 2/3$  bone affected. The following rules were agreed for MTP joints: the proximal and distal bones were scored separately and summed for a total sclerosis score (maximum score 6 for each joint), the bone within 1-cm distance from the articular surface was scored, and if the bone was eroded, available bone was used as the denominator. Only the anterior-posterior view was used to assess sclerosis.

A key challenge for spur scoring was that joints within the gout scoring system have variable size and affected bones could have more than one spur. For this reason, the Thumb Base Osteoarthritis Magnetic Resonance Imaging Scoring System [10] was modified for scoring spurs in gout; this system captured both spur size and number: 0

= no spurs, 1 = 1 small spur, 2 =  $\geq 2$  small spurs and/or 1 moderate spur, and 3 =  $\geq 2$  moderate spurs and/or  $\geq 1$  large spur(s). As with sclerosis, the proximal and distal bones were scored for spur separately and summed for a total spur score, and the bone within a 1-cm distance from the articular surface was scored. Both the anterior-posterior and oblique views were used to assess spurs.

The standardized definitions and projections are shown in Table 1, and examples of the scoring system are shown in Fig. 1. Rescoring of 80 individual joints (40 1st MTP joints and 40 5th MTP joints) for sclerosis and spur using this new scoring system and associated rules led to an improved inter-reader agreement: for sclerosis 0.84 (0.76–0.89) and for spur 0.81 (0.72–0.87) (Table 2).

#### Comparison with CT measurements in individual joints

To assess construct validity, we compared the plain radiography scores with CT measurements of sclerosis and spur in individual joints. For sclerosis, the plain radiographic sclerosis score correlated with CT sclerosis ratio ( $r = 0.71$ ,  $P < 0.001$ ) and Hounsfield units ( $r = 0.74$ ,  $P < 0.001$ ). For joints with sclerosis on CT, the median (IQR) plain radiographic sclerosis score was 2 (1–2), compared with 0 (0–0) in joints without sclerosis ( $P < 0.001$ ) (Table 3). For spur, the plain radiographic spur score correlated with CT spur length in millimeters ( $r = 0.65$ ,  $P < 0.001$ ), and spur ratio ( $r = 0.65$ ,  $P < 0.001$ ). For joints with spur on CT, the median (IQR) plain radiographic spur score was 1 (1–2), compared with 0 (0–0) in joints without spur ( $P < 0.001$ ) (Table 3).

#### Testing of the scoring system in sets of hand and foot radiographs

The scoring system was then tested in sets of hand and foot radiographs from 25 people with gout, obtained at baseline and after 2 years. Over the 2-year period, no significant differences were observed in sclerosis scores which changed from median (IQR) 8.5 (6–25) at baseline to 10 (5–25) at year 2,  $P = 0.076$ , or spur scores which changed from median (IQR) 4 (0–9) to 6 (1–8),  $P = 0.70$  (Table 4). However, median (IQR) erosion scores increased from 24 (13–64) to 28 (18–70),  $P = 0.001$ .

The inter-reader ICCs for the baseline radiographs were 0.94 (0.90–0.98) for sclerosis and 0.76 (0.65–0.84) for spur (Table 2). For the baseline radiographs, plain radiographic erosion scores correlated with sclerosis scores (for average of both readers,  $r = 0.87$ ,  $P < 0.001$ ) and spur scores ( $r = 0.71$ ,  $P < 0.001$ ) (Table 5). Sclerosis and spur scores also correlated (for average of both readers,  $r = 0.66$ ,  $P < 0.001$ ). Change in sclerosis and change in spur scores were correlated ( $r = 0.71$ ,  $P < 0.001$ ). However, change in erosion scores did not

**Table 1** Radiographic scoring system for new bone formation in gout

<b>Sclerosis</b>	
<b>Definition</b>	Increased density of medullary or subcortical bone.
<b>Scoring rules</b>	The proximal and distal bones are scored separately for interphalangeal joints, metacarpophalangeal joints, and metatarsophalangeal joints (maximum score for individual joint is 6). Single bones are scored for the carpal and tarsal articulations (maximum score for individual bone is 3). For long bones, score the 1-cm distance from the articular surface. If the bone is eroded, score the available bone as the denominator.
<b>Joints scored</b>	The same bones scored in the gout-modified Sharp-van der Heijde scoring system for bone erosion (hand distal interphalangeal joints, hand proximal interphalangeal joints, interphalangeal joint of the thumb, metacarpophalangeal joints, base of the first metacarpal, multangular, scaphoid, lunate, radius, ulna, metatarsophalangeal joints, and interphalangeal joints of the hallux).
<b>Series and views</b>	Hand: anterior-posterior. Foot: anterior-posterior.
<b>Score</b>	0 = no sclerosis. 1 = < 1/3 bone affected. 2 = 1/3-2/3 bone affected. 3 = > 2/3 bone affected.
<b>Total score range</b>	0–276.
<b>Spur</b>	
<b>Definition</b>	A sharp spicule of dense bone proliferation extending at an acute angle from the cortex. This should be differentiated from an osteophyte, which arises along the joint margin.
<b>Scoring rules</b>	The proximal and distal bones are scored separately for interphalangeal joints, metacarpophalangeal joints, and metatarsophalangeal joints (maximum score for individual joint is 6). Single bones are scored for the carpal and tarsal articulations (maximum score for individual bone is 3). For long bones, score the 1-cm distance from the articular surface.
<b>Rules</b>	The same bones scored in the gout-modified Sharp-van der Heijde scoring system for bone erosion (hand distal interphalangeal joints, hand proximal interphalangeal joints, interphalangeal joint of the thumb, metacarpophalangeal joints, base of the first metacarpal, multangular, scaphoid, lunate, radius, ulna, metatarsophalangeal joints, and interphalangeal joints of the hallux).
<b>Series and views</b>	Hand: anterior-posterior and ball-catcher. Foot: anterior-posterior and oblique.
<b>Score</b>	0 = no spurs. 1 = 1 small spur. 2 = ≥ 2 small spurs and/or 1 moderate spur. 3 = ≥ 2 moderate spurs and/or ≥ 1 large spur(s).
<b>Total score range</b>	0–276.

**Table 2** Inter-reader intraclass correlation coefficient (ICC) for the individual joints and hand and foot radiograph sets. For hand and foot radiograph sets, ICCs for baseline visit are shown

New bone formation feature	Individual joint analysis (n = 80 joints)	Hand and foot radiograph sets (n = 25 patients)
Sclerosis	0.84 (0.76–0.89)	0.94 (0.90–0.98)
Spur	0.81 (0.72–0.87)	0.76 (0.65–0.84)

correlate with change in sclerosis scores ( $r = -0.04$ ,  $P = 0.86$ ) or change in spur scores ( $r = 0.15$ ,  $P = 0.48$ ).

At the baseline visit, there was no association between new bone formation scores and gout disease duration or gout flare frequency (Supplementary Table 6). Spur scores correlated with serum urate at baseline ( $r = 0.41$ ,  $P = 0.04$ ), and change in spur score correlated with change in serum urate over the 2-year study period ( $r = 0.41$ ,  $P = 0.04$ ) (Supplementary Table 7).

**Table 3** Plain radiographic scores in individual joints according to computed tomography (CT) reference standard. Average plain radiographic scores for both readers were used in this analysis. Mann-Whitney *U* test *P* values are shown

	Number of joints	Median (IQR) plain radiographic sclerosis score	<i>P</i>
CT sclerosis present	34	2 (1–2)	< 0.001
CT sclerosis absent	46	0 (0–0)	
	Number of joints	Median (IQR) plain radiographic spur score	
CT spur present	19	1 (0–2)	< 0.001
CT spur absent	61	0 (0–0)	

IQR interquartile range

**Discussion**

This project has identified a feasible and reproducible method for scoring new bone formation features of sclerosis and spur in gout. This method correlates well with

**Table 4** Change over time in radiographic parameters. Average plain radiographic scores for both readers were used in this analysis. Wilcoxon signed-rank test *P* values are shown

	Baseline	Year 2	<i>P</i>
<b>Sclerosis score, median (IQR)</b>	8.5 (5.5–25)	10 (5.25–8.25)	0.076
<b>Spur score, median (IQR)</b>	4.0 (0.25–8.5)	5.5 (0.75–8.25)	0.70
<b>Erosion score, median (IQR)</b>	24 (13–64)	28 (17.5–70)	0.001

*IQR* interquartile range

reference standard CT measurement and has good inter-reader reproducibility. This scoring system may facilitate research examining the mechanisms, impact, and treatment of structural bone lesions in gout, beyond bone erosion, which, with a few exceptions [1, 14], has been the major focus of research to date [15–18].

The literature review that informed the development of this scoring system showed substantial heterogeneity in radiological scoring systems for features of new bone formation both within and between different musculoskeletal diseases. In gout, some features of new bone formation are unique, and no published scoring system was easily transferable to gout. The initial scoring exercises demonstrated many of the challenges of scoring new bone formation and the need for clear definitions, standardization of projections, and reference images.

No previous plain radiographic new bone formation scoring systems have used a purely quantitative score. When this approach was tested for gout, inter-reader reliability was low, and this approach was also time-consuming. The time required for measurements and reliability may improve with the use of machine learning which has been demonstrated to reliably detect radiographic progression in axial spondyloarthritis [19]. The project also showed the limitations of a single measurement of gout new bone formation features such as the proliferation score used in the PARS system; this method lacked inter-reader reproducibility when used in gout, and had low face validity for gout, as the features of new bone formation in gout are diverse and may not change in a consistent way. For this reason, we did not progress

a single or composite scoring system following the initial scoring exercise.

Consistent with prior work [1, 3], this study suggests complex links between new bone formation and bone erosion in gout. The tophus has been strongly implicated in the development of bone erosion in gout [20–24], and prior research has shown associations between erosion, tophus, and new bone formation in gout [1], indicating that new bone formation may be repair phenomena triggered by the bone damage induced by the tophus. While the cross-sectional correlation analysis between erosion, spur, and sclerosis scores suggests that these lesions may reflect a generalized joint pathology, it is noteworthy that in the longitudinal analysis, change in erosion scores did not correlate with change in spur scores or change in sclerosis scores. These findings suggest that bone erosion and proliferative changes in gout may not be completely linked and that new bone formation scores may provide additional value to the assessment of structural bone disease in gout.

It is noteworthy that although there were small increases in bone erosion in the serial radiographs using the gout-modified Sharp-van der Heijde erosion score, no significant change in sclerosis or spur scores were observed over 2 years. These findings raise some uncertainty about the sensitivity to change of this scoring system, particularly noting that prior descriptive series have reported increased bone sclerosis within the affected joints in response to intensive urate-lowering therapy such as pegloticase [3, 4]. Importantly, participants in the longitudinal analysis were already established on urate-lowering therapy at the baseline assessment. Additional testing of the scoring system following initiation of urate-lowering therapy, following intensive urate-lowering, and particularly using randomized controlled clinical trial databases will be important to further understand the properties and performance of the scoring system. The future research agenda will also include testing whether new bone formation influences joint function, understanding the relationship between NBF and erosion healing in response to urate-lowering therapy, and whether other treatments such as anti-osteoclast therapy with

**Table 5** Spearman correlation coefficients for plain radiographic sclerosis, spur, and erosion total scores at baseline visit (*n* = 25) and for change over 2 years. Average plain radiographic scores for both readers were used in this analysis

		Baseline score		Year 2 scores		Change in scores	
		Sclerosis	Spur	Sclerosis	Spur	Sclerosis	Spur
<b>Erosion score</b>	<i>r</i>	0.87	0.71	0.88	0.68	−0.04	0.15
	<i>P</i>	< 0.001	< 0.001	< 0.001	< 0.001	0.86	0.48
<b>Sclerosis score</b>	<i>r</i>	—	0.66	—	0.66	—	0.71
	<i>P</i>	—	< 0.001	—	< 0.001	—	< 0.001

zoledronate or denosumab influence new bone formation despite the lack of effect on erosion and joint space narrowing [16, 25]. Although we did observe an association between serum urate and spur scores over time, the study design did not allow us to determine the relationship between new bone formation scores and urate or inflammatory burden over the entire period of the disease (noting the mean disease duration was approximately 20 years); future studies should assess the cumulative burden of hyperuricemia and gout flares on new bone formation.

## Conclusion

In conclusion, a semi-quantitative plain radiographic scoring method for the assessment of bone sclerosis and spur in gout is feasible, valid, and reproducible. This method may facilitate consistent measurement of new bone formation in gout.

## Abbreviations

CI: Confidence intervals; CT: Computed tomography; ICC: Intraclass correlation coefficients; IP: Interphalangeal; IQR: Interquartile range; MSK: Musculoskeletal; MTP: Metatarsophalangeal; NBF: New bone formation; PARS: Proliferation score of the Psoriatic Arthritis Ratingen Score; SD: Standard deviation; TOMS: Thumb Base Osteoarthritis Magnetic Resonance Imaging Scoring System.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13075-021-02683-9>.

**Additional file 1: Figure S1.** PRISMA flow diagram for literature review. **Table S1.** Frequency of the most commonly used scoring systems that assess any feature of new bone formation ( $n = 840$  papers)<sup>a</sup>. Data are presented for the most frequently used radiographic scoring systems (> 5 articles). **Table S2.** Measurements for the first development exercise of individual joint analysis scoring. **Table S3.** Clinical features of the 20 participants in development exercise. Unless stated, data are presented as mean (SD). **Table S4.** Clinical features of the 25 participants with full scoring exercise. Unless stated, data are presented as mean (SD). **Table S5.** Intraclass correlation coefficient (ICC) for the first development exercise. ICC, intraclass correlation coefficient; CI, confidence interval; BLOKS, Boston Leeds Osteoarthritis Knee Score; PARS, Psoriatic Arthritis Ratingen Score. **Table S6.** Spearman correlation coefficients for baseline visit gout characteristics and new bone formation scores in the full scoring exercise. **Table S7.** Spearman correlation coefficients for changes in gout outcomes and new bone formation scores in the full scoring exercise over the Year 2 study period. Data are shown as Spearman correlation coefficient (P value).

## Acknowledgements

None.

## Authors' contributions

CS, KC, AD, and ND were responsible for the conception and design. CS, KC, SS, JF, KB, YJT, AD, and ND contributed to the data extraction and interpretation. CS, KC, JF, KB, TB, AH, LS, AD, and ND contributed to the design and conduction of statistical analysis. CS, KC, SS, and ND prepared the manuscript. JF, KB, YJT, TB, AH, LS, AD, and ND supervised the manuscript. All authors read and approved the final manuscript.

## Funding

Changnam Son was supported by the National Research Foundation of Korea [grant number NRF-2021R1F1A1060970 and NRF-2018M3C1B7020722]. Ken Cai was supported by Arthritis Australia [Leanne Stafford Award]. The Health Research Council of New Zealand supported this work (11-203 and 15-576).

## Availability of data and materials

The data that supported the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

## Declarations

### Ethics approval and consent to participate

Plain radiographs and computed tomography (CT) scans were obtained for research, approved by the New Zealand Ministry of Health (NTX/08/06/050/AM06 and MEC/11/06/060). All participants provided written informed consent.

### Consent for publication

Not applicable

### Competing interests

ND reports grants and personal fees from AstraZeneca; personal fees from PK Med, JW Pharmaceutical Corporation, Horizon, Abbvie, AstraZeneca, Janssen, Dyve Biosciences, Selecta, and ArthroSi; and grants from Amgen and AstraZeneca outside the submitted work. The other authors declare that they have no competing interests.

### Author details

<sup>1</sup>Keimyung University School of Medicine, Daegu, South Korea. <sup>2</sup>Department of Medicine, Faculty of Medical and Health Sciences, The University of Auckland, Auckland, New Zealand. <sup>3</sup>Department of Radiology, Auckland District Health Board, Auckland, New Zealand. <sup>4</sup>Department of Rheumatology, Hôpital Lariboisière, Paris, France. <sup>5</sup>Department of Medicine, University of Otago Christchurch, Christchurch, New Zealand. <sup>6</sup>Department of Radiology with Anatomy, Faculty of Medical and Health Sciences, The University of Auckland, Auckland, New Zealand.

Received: 1 October 2021 Accepted: 16 November 2021

Published online: 08 December 2021

## References

- Dalbeth N, Milligan A, Doyle AJ, Clark B, McQueen FM. Characterization of new bone formation in gout: a quantitative site-by-site analysis using plain radiography and computed tomography. *Arthritis Res Ther*. 2012;14:R165.
- Dalbeth N, Clark B, McQueen F, Doyle A, Taylor W. Validation of a radiographic damage index in chronic gout. *Arthritis Rheum*. 2007;57:1067–73.
- Dalbeth N, Doyle A, McQueen FM. Clinical images: divergent patterns of joint remodeling following effective urate-lowering therapy in tophaceous gout. *Arthritis Rheum*. 2011;63:266.
- Dalbeth N, Doyle AJ, McQueen FM, Sundry J, Baraf HS. Exploratory study of radiographic change in patients with tophaceous gout treated with intensive urate-lowering therapy. *Arthritis Care Res*. 2014;66:82–5.
- Neogi T, Jansen TL, Dalbeth N, Fransen J, Schumacher HR, Berendsen D, et al. 2015 Gout classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Ann Rheum Dis*. 2015;74:1789–98.
- Haugen IK, Lillegraven S, Slatkowsky-Christensen B, Haavardsholm EA, Sesseng S, Kvien TK, et al. Hand osteoarthritis and MRI: development and first validation step of the proposed Oslo Hand Osteoarthritis MRI score. *Ann Rheum Dis*. 2011;70:1033–8.
- Hunter DJ, Lo GH, Gale D, Grainger AJ, Guermazi A, Conaghan PG. The reliability of a new scoring system for knee osteoarthritis MRI and the validity of bone marrow lesion assessment: BLOKS (Boston Leeds Osteoarthritis Knee Score). *Ann Rheum Dis*. 2008;67:206–11.

8. Wassenberg S, Fischer-Kahle V, Herborn G, Rau R. A method to score radiographic change in psoriatic arthritis. *Z Rheumatol.* 2001;60:156–66.
9. Kannus P, Jarvinen M, Paakkala T. A radiological scoring scale for evaluation of post-traumatic osteoarthritis after knee ligament injuries. *Int Orthop.* 1988;12:291–7.
10. Kroon FPB, Conaghan PG, Foltz V, Gandjbakhch F, Peterfy C, Eshed I, et al. Development and reliability of the OMERACT thumb base osteoarthritis magnetic resonance imaging scoring system. *J Rheumatol.* 2017;44:1694–8.
11. Poggenborg RP, Bird P, Boonen A, Wiell C, Pedersen SJ, Sørensen IJ, et al. Pattern of bone erosion and bone proliferation in psoriatic arthritis hands: a high-resolution computed tomography and radiography follow-up study during adalimumab therapy. *Scand J Rheumatol.* 2014;43:202–8.
12. Döhn UM, Ejbjerg BJ, Hasselquist M, Narvestad E, Møller J, Thomsen HS, et al. Detection of bone erosions in rheumatoid arthritis wrist joints with magnetic resonance imaging, computed tomography and radiography. *Arthr Res Ther.* 2008;10:R25.
13. van der Heijde D. How to read radiographs according to the Sharp/van der Heijde method. *J Rheumatol.* 2000;27:261–3.
14. Pecherstorfer C, Simon D, Unbehend S, Ellmann H, Englbrecht M, Hartmann F, et al. A detailed analysis of the association between urate deposition and erosions and osteophytes in gout. *ACR Open Rheumatol.* 2020;2:565–72.
15. Stewart S, Aati O, Horne A, Doyle AJ, Dalbeth N. Radiographic damage scores predict grip strength in patients with tophaceous gout. *Rheumatology (Oxford).* 2020;59:1440–2.
16. Dalbeth N, Aati O, Gamble GD, Horne A, House ME, Roger M, et al. Zoledronate for prevention of bone erosion in tophaceous gout: a randomised, double-blind, placebo-controlled trial. *Ann Rheum Dis.* 2014;73:1044–51.
17. Dalbeth N, Billington K, Doyle A, Frampton C, Tan P, Aati O, et al. Effects of allopurinol dose escalation on bone erosion and urate volume in gout: a dual-energy computed tomography imaging study within a randomized, controlled trial. *Arthritis Rheum.* 2019;71:1739–46.
18. Dalbeth N, Saag KG, Palmer WE, Choi HK, Hunt B, MacDonald PA, et al. Effects of febuxostat in early gout: a randomized, double-blind, placebo-controlled study. *Arthritis Rheum.* 2017;69:2386–95.
19. Joo YB, Baek IW, Park YJ, Park KS, Kim KJ. Machine learning-based prediction of radiographic progression in patients with axial spondyloarthritis. *Clin Rheumatol.* 2020;39:983–91.
20. Dalbeth N, Aati O, Kalluru R, Gamble GD, Horne A, Doyle AJ, et al. Relationship between structural joint damage and urate deposition in gout: a plain radiography and dual-energy CT study. *Ann Rheum Dis.* 2015;74:1030–6.
21. McQueen FM, Doyle A, Reeves Q, Gao A, Tsai A, Gamble GD, et al. Bone erosions in patients with chronic gouty arthropathy are associated with tophi but not bone oedema or synovitis: new insights from a 3 T MRI study. *Rheumatology (Oxford).* 2014;53:95–103.
22. Towiwat P, Doyle AJ, Gamble GD, Tan P, Aati O, Horne A, et al. Urate crystal deposition and bone erosion in gout: 'inside-out' or 'outside-in'? A dual-energy computed tomography study. *Arthritis Res Ther.* 2016;18:208.
23. Sapsford M, Gamble GD, Aati O, Knight J, Horne A, Doyle AJ, et al. Relationship of bone erosion with the urate and soft tissue components of the tophus in gout: a dual energy computed tomography study. *Rheumatology (Oxford).* 2017;56:129–33.
24. Shi D, Chen JY, Wu HX, Zhou QJ, Chen HY, Lu YF, et al. Relationship between urate within tophus and bone erosion according to the anatomic location of urate deposition in gout: a quantitative analysis using dual-energy CT volume measurements. *Medicine (Baltimore).* 2019;98:e18431.
25. Gaffo AL, Saag K, Doyle AJ, Melnick J, Horne A, Foster J, et al. Denosumab did not improve computerized tomography erosion scores when added to intensive urate-lowering therapy in gout: results from a pilot randomized controlled trial. *Semin Arthritis Rheum.* 2021;51:1218–23.

## Publisher's Note

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

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

