Article:
Sirotti, S., Terslev, L., Filippucci, E. et al. (2023) Development and validation of an OMERACT ultrasound scoring system for the extent of calcium pyrophosphate crystal deposition at the joint level and patient level. The Lancet Rheumatology, 5 (8). e474-e482. ISSN: 2665-9913

https://doi.org/10.1016/s2665-9913(23)00136-4

Title: Development and Validation of an OMERACT Ultrasound Scoring System for the Extent of Calcium Pyrophosphate Crystal Deposition at Joint and Patient Level

Authors:

1. Silvia Sirotti, MD, IRCCS Ospedale Galeazzi – Sant'Ambrogio, Rheumatology Department, Milan, Italy, silvia.sirotti@gmail.com
2. Lene Terslev, Full Professor, Center for Rheumatology and Spine Diseases, Rigshospitalet, Copenhagen, Denmark AND Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark, terslev@dadlnet.dk
3. Emilio Filippucci, Associate Professor, Rheumatology Unit, Department of Clinical and Molecular Sciences, Polytechnic University of Marche, Ancona, Italy, emilio_filippucci@yahoo.it
4. Annamaria Iagnocco, Full Professor, Department of Clinical and Biological Sciences, Università degli Studi di Torino, Turin, Italy, annamaria.iagnocco1@gmail.com
5. Ingrid Moller, MD, Instituto Poal de Reumatologia, University of Barcelona, Barcelona, Spain, ingrid.moller@ipoal.com
6. Esperanza Naredo, Associate Professor, Rheumatology Department, Joint and Bone Research Unit, Hospital Universitario Fundación Jiménez Díaz and IIS-FJD, Universidad Autónoma of Madrid, Madrid, Spain, enaredo@ser.es
7. Florentin A. Vreju, Full Professor, Rheumatology Department, University of Medicine and Pharmacy of Craiova, Craiova, Romania, florin_vreju@yahoo.com
8. Antonella Adinolfi, MD, Rheumatology Division, ASST Grande Ospedale Metropolitano Niguarda, Milan, Italy, antonella.adinolfi986@gmail.com
9. Fabio Becce, Associate Professor, Department of Diagnostic and Interventional Radiology, Lausanne University Hospital, University of Lausanne, Lausanne, Switzerland, fabio.becce@chuv.ch
10. Hilde Berner Hammer, Full Professor, Center for treatment of Rheumatic and Musculoskeletal Diseases (REMEDY), Diakonhjemmet Hospital, Oslo, Norway AND Faculty of Medicine, University of Oslo, Oslo, Norway, hbham@online.no
11. Tomas Cazenave, MD, Rheumatology Unit, Instituto de Rehabilitación Psicofísica, Buenos Aires, Argentina, tomascazenave@hotmail.com

12. Edoardo Cipolletta, MD, Rheumatology Unit, Department of Clinical and Molecular Sciences, Polytechnic University of Marche, Ancona, Italy, edoardocipolletta@gmail.com
13. Sara Nysom Christiansen, MD, Center for Rheumatology and Spine Diseases, Rigshospitalet, Copenhagen, Denmark, sara.nysom.christiansen@regionh.dk
14. Andrea Delle Sedie, MD, Rheumatology Unit, Azienda Ospedaliero-Universitaria Pisana, Pisa, Italy, adellesedie@gmail.com
15. Mario Diaz, MD, Rheumatology Unit, University Hospital Fundación Santa Fe de Bogota, Bogota, Colombia, diazmachi@hotmail.com
16. Fabiana Figus, MD, Rheumatology Service, ASL Torino 3, Collegno Pinerolo, Turin, Italy, fabianafigus@gmail.com
17. Peter Mandl, Associate Professor, Division of Rheumatology, Department of Internal Medicine III, Medical University Vienna, Vienna, Austria, dr.peter.mandl@gmail.com
18. Daryl MacCarter, MD, Department of Rheumatology, North Valley Hospital, Whitefish, Montana, USA, dkmaccarter@gmail.com
19. Mohamed A. Mortada, Full Professor, Rheumatology Department, Faculty of Medicine, Zagazig University, Zagazig, Egypt, m_a_mortada@yahoo.com
20. Gael Mouterde, MD, Rheumatology Department & IDESP, CHU Montpellier, Montpellier University, Montpellier, France, g-mouterde@chu-montpellier.fr
21. Francesco Porta, MD, Interdisciplinary Pain Medicine Unit, Rheumatology Section, Santa Maria Maddalena Hospital, Italy, dr.fporta@gmail.com
22. Anthony M. Reginato, Full Professor, Division of Rheumatology, The Warren Alpert Medical School of Brown University, Providence, RI, USA, anthonyreginato@yahoo.com
23. Wolfgang A. Schmidt, Full Professor, Immanuel Krankenhaus Berlin, Medical Centre for Rheumatology Berlin-Buch, Berlin, Germany, w.schmidt@immanuel.de
24. Teodora Serban, MD, Rheumatology Department, Ospedale La Colletta, ASL3 Genovese, Genoa, Italy, teodoramazdrag@yahoo.com
25. Richard J. Wakefield, MD, Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds AND Leeds Biomedical Research Centre, Leeds Teaching Hospitals Trust, Chapel Allerton Hospital, Leeds, UK, R.J.Wakefield@leeds.ac.uk
26. Pascal Zufferey, MD, Rheumatology Department, University of Lausanne, CHUV, Lausanne, Switzerland, zuffereypascal@gmail.com

27. Piercarlo Sarzi-Puttini, Full Professor, IRCCS Ospedale Galeazzi – Sant'Ambrogio, Rheumatology Department, Milan, Italy AND Department of Biomedical and Clinical Sciences, Università degli Studi di Milano, Milan, Italy, piercarlo.sarziputtini@gmail.com
28. Anna Zanetti, PhD, Società Italiana di Reumatologia, Epidemiology Research Unit, Milan, Italy, a.zanetti@reumatologia.it
29. Arianna Damiani, MD, Department of Experimental and Clinical Medicine, University of Florence, Florence, Italy, ariannadam@outlook.it
30. Carlos Pineda, MD, Division of Rheumatology, Instituto Nacional de Rehabilitacion Luis Guillermo Ibarra Ibarra, Mexico City, Mexico, carpineda@yahoo.com
31. Helen I. Keen, Associate Professor, School of Medicine, The University of Western Australia, Murdoch, Perth, Australia, helen.keen@uwa.edu.au
32. Maria Antonietta D'Agostino, Full Professor, Rheumatology Department, Università Cattolica del Sacro Cuore, Fondazione Policlinico Universitario Agostino Gemelli IRCSS, Rome, Italy, mariaantonietta.dagostino@unicatt.it
33. Georgios Filippou, MD, IRCCS Ospedale Galeazzi – Sant'Ambrogio, Rheumatology Department, Milan, Italy, gf.filippou@gmail.com

Corresponding author: Georgios Filippou MD, Rheumatology Department, IRCCS Ospedale Galeazzi – Sant'Ambrogio, Via Cristina Belgioioso 173, 20157 Milan, Italy; office phone: +390283500010; e-mail: gf.filippou@gmail.com.

Research in context:

Evidence before this study

Calcium pyrophosphate deposition (CPPD) is still a neglected disease even if appears quite prevalent in the elderly. However, in recent years, awareness of CPPD is increasing, thanks to the effort made by international task forces that are collaborating to create classification criteria, recommendations for the use of imaging and validation of outcome measures. Nevertheless, there are currently no validated instruments available for monitoring CPPD patients over time. Ultrasound (US) has proven to be an accurate technique for diagnosing CPPD and the OMERACT US working group - CPPD

subgroup was founded to develop and validate a US-related outcome measurement instrument in CPPD and has recently developed and validated ultrasonographic definitions for CPPD.

We searched PubMed database for publications in English from 1962, when the disease was first described, until Apr 2021, regarding US scoring for the assessment of the extent of CPPD in the literature. The only attempt was made in 2013 by Filippou et al, but the scoring system was never validated.

Added value of this study

This is the first study to develop and validate a standardized scoring system for CPPD extent, based on consensus among experts. The validation process for using US as an outcome measurement instrument follows the rigorous OMERACT US methodology, which entails a stringent step-by-step roadmap, beginning with a systematic literature review conducted according to the OMERACT Framework Instrument Selection Algorithm (OFISA) recommendations; subsequently, the process included a Delphi consensus procedure, as well as image- and patient-based reliability studies. The project has facilitated the collaboration of leading experts in the field of US and CPPD, resulting in the creation of a scoring system based on widely standardized methodology.

Implications of all the available evidence

This study represents a crucial step towards the validation of US as an outcome measurement instrument in CPPD, leading to greater homogeneity between clinical trials and ensuring valid and comparable results in CPPD research. Additionally, this study provides a valuable tool for monitoring the extent of crystal deposition in patients with CPPD in clinical practice, which could offer relevant insights into the pathogenesis and natural history of this neglected disease.

Keywords*:* Chondrocalcinosis; Crystal arthropathies; Outcome Measure Instrument; Ultrasonography.

Abstract

Background: The OMERACT Ultrasound (US) working group - Calcium Pyrophosphate Deposition (CPPD) subgroup has recently developed and validated the elementary lesions for the detection of calcium pyrophosphate crystals in joints. The aim of this step was to develop and evaluate the reliability of a consensus-based US scoring system for CPPD extent.

Methods: According to the OMERACT US stepwise process, the scoring system was developed through a Delphi survey based on a systematic literature review and expert opinion. The reliability of the score was assessed through exercises on static images and on patients.

Findings: The four-grade semiquantitative scoring system consisted of: grade 0: no findings consistent with CPPD; grade 1: ≤3 single spots or 1 small deposit; grade 2: > 3 single spots or >1 small deposit or ≥1 larger deposit occupying ≤ 50% of the structure under examination in the reference image (i.e., the scanning view with the highest grade of depositions); grade 3: deposits that occupy more than 50% of the structure under examination in the reference image. The score should be applied to the knee (menisci and hyaline cartilage) and the triangular fibrocartilage complex of the wrist. The intra- and inter-reader reliabilities in static images were almost perfect (kappa 0.90 and 0.84, respectively), and the reliabilities on patients were substantial (kappa 0.72 for intra-reader and 0.66 for inter-reader).

Interpretations: This OMERACT US scoring system for CPPD was reliable on both static images and patients. It may be a valuable tool for ensuring valid and comparable results in clinical trials, and it could help monitor the extent of crystal deposition in patients with CPPD in clinical practice.

Funding: This paper was supported and funded by the Italian Ministry of Health - "Ricerca Corrente".

Introduction

Calcium pyrophosphate deposition (CPPD) is defined as "the umbrella term used to describe all instances of calcium pyrophosphate (CPP) crystal deposition in tissues"[1]. It is considered to be a common condition among the elderly[2], with an estimated prevalence ranging from 7% to 13.7% depending on the age of population, the joint assessed, and the diagnostic criteria used[3]. Despite the first description of CPPD disease stemming back to 1962[4], there are still major shortfalls in its nomenclature, pathogenesis, diagnosis, outcome measurement instruments, and treatment[5]. All these unmet needs have had and still have a significant impact on both clinical research and management of CPPD.

Over the last decades, musculoskeletal ultrasound (US) has progressively gained a key role in rheumatological clinical practice and has proven to be an excellent tool for detecting CPP deposits[6,7]. US application in CPPD was first described in 1990 by Kellner and colleagues[8], and since then a growing number of studies have confirmed its diagnostic value[9,10]. However, the significant heterogenicity among studies about US definitions used, the assessed joints and the adopted reference standard, has hampered the comparability of study results. To standardise US research in this disease, the Outcome Measures in Rheumatology (OMERACT) US working group (WG) - CPPD subgroup was established in 2014 with the aim of validating US as an outcome measurement instrument for CPPD. Following a stepwise process[11], the group developed a set of definitions for ultrasonographic findings[12], which were demonstrated to be reliable and accurate for CPPD diagnosis[13,14]. The next step in the OMERACT stepwise approach was to develop and validate a new scoring system that allowed for the quantification and extent of CPP crystal deposition in patients, which could be a useful tool for CPPD assessment over time. To date, no validated scoring system has been published in the literature, and the only attempt to generate one, carried out in 2013 by Filippou *et al*.[15], was evaluated only in two centers and was not adequately validated.

The aims of this study were to develop a consensus-based US scoring system for CPPD extent, which quantifies the area of targeted tissues covered by CPP crystals deposition both at joint and patient level, and to evaluate its reliability in static images and on patients, according to the OMERACT US methodology. This will ensure valid and comparable results in clinical trials and facilitate the monitoring of crystal deposition extent in clinical practice.

Methods

The novel scoring system was developed through a stepwise process, following the established OMERACT US methodology[11]. Firstly, a systematic literature review (SLR) was conducted to gather available evidence on existing scoring system for CPPD and to estimate the prevalence of CPPD in peripheral joints using imaging[16]. The group then focused on developing the scoring system through a Delphi exercise. Finally, the reliability of the proposed scoring system was tested in a web-based and subsequently in a patient-based exercise.

Preliminary survey and Delphi exercise

Prior to conducting the Delphi survey, a preliminary survey was circulated among the members of the OMERACT US WG - CPPD subgroup to collect their suggestions on the features that should be included in the scoring system according to their personal experience.

With the aim to achieve a consensus on the items and sites to include in the scoring system, a Delphi questionnaire was prepared in the REDCap platform. The questionnaire reflected the results of the literature review and the expert opinion of the steering group and was sent to 41 members of the group.

Participants rated each statement on a 5-point Likert scale, which was graded as follows: 1, strongly disagree; 2, disagree; 3, neither agree nor disagree; 4, agree; 5, strongly agree. Statements that received at least 75% of responses rated as 4 and 5 were considered as appropriate for inclusion in the final scoring system. In case of disagreement, statements that received 4 or 5 rating from less than 50% of respondents were excluded. The remaining statements were presented for voting in subsequent rounds after being modified according to members' suggestions and comments, until agreement was reached for all statements. To keep track of the evolution of the scoring system, the agreed and disagreed statements from previous rounds were summarized and displayed with the mean scores in the following round by the organizers (G.F. and S.S.).

Reliability exercises

After developing the new US scoring system for CPPD extent, the OMERACT US process foresaw its validation on static images and on patients. The first step involved a web-based exercise, where all

group members were invited to participate. After the Delphi survey, participants were asked to collaborate in the creation of an image collection that would be representative of the developed scoring system. The collected images were evaluated for quality; and finally, a set of images that equally represented all degrees of crystal deposition in the selected sites according to the scoring system was defined. The images were inserted into a survey in the REDCap platform in random order, and two rounds of a web-based exercise were conducted with an interval of 4 weeks. The first round aimed to assess the inter-reader reliability, and the second to assess the intra-reader reliability of the new scoring system on static images. All participants voted for each US image according to the new scoring system. The atlas with representative images of each score was visible throughout the entire exercise to facilitate the application of the definitions and to avoid scoring according to personal experience or opinion.

In the second step of the validation process, a patient-based exercise was conducted with the participation of 8 experts of the OMERACT US WG - CPPD subgroup (M.A.D'A., G.F., E.F., A.I., I.M., E.N., L.T., F.V). All of them have more than 15 years' experience in US and are members of the EULAR faculty for US training. The workshop took place in Milan, Italy, and patients with crystal-proven CPPD (with positive synovial fluid analysis) were recruited from the Rheumatology Department of Luigi Sacco University Hospital. Eight high-end US machines (4 Samsung RS85 prestige, 1 Samsung RS80 prestige and 3 Samsung V8) equipped with a 2-14 MHz linear probe were used for the workshop. All machines were equipped with the same high resolution linear probes, and the same settings. The sonographers were free to modify the basic functions, such as depth, gain, time gain compensation (TGC) and frequency, to achieve optimal images for CPPD identification. Power Doppler examination was used upon sonographers' judgment to avoid any pitfalls or artifacts (i.e., identification of small vessels). A short calibration briefing was conducted before the exercise.

The scanning protocol was determined after the development of the scoring system and included bilateral evaluation of the triangular fibrocartilage complex (TFCC) of the wrist, knee menisci and knee hyaline cartilage (HC). After calculating the sample size (see statistical analysis paragraph), 8 patients were recruited, and US exams were carried out twice on each patient on the same day by the 8 experts using the newly developed scoring system, to assess the intra- and inter-reader reliability on patients. Each sonographer was given 12 minutes to assess the target joints, after which they moved to the next patient until all patients were examined. The patients' locations were

randomly changed for the second round. The atlas with representative images was visible throughout the exercise.

The US exams were performed according to a standardized sequence, using techniques already described in the literature for the identification of CPPD deposits[13,17]. Briefly, the knee menisci were examined, using both transverse and longitudinal scans[18], with the knee in complete extension and in semi-flexed position (30-45°) sliding with the probe (without lifting it) over the structure under examination from the anterior to the posterior horn. Knee HC was scanned with the knee in maximum flexion for the anterior portion with both transverse and longitudinal scans. The posterior portion of the HC was ignored to avoid patients to roll over several times as requested by the protocol. The TFCC of the wrist was examined by sliding the probe over the structure, without lifting it, from the dorsal to the palmar aspect for longitudinal scanning and from proximal to distal for the transverse scanning. In all cases, dynamic scanning was allowed if considered necessary, such as flexion-extension of the knee or radial deviation of the wrist.

This study was approved by the institutional ethics committee of Luigi Sacco University Hospital, Milan, Italy (Registro Sperimentazioni n. 2022/ST/145). All participants provided written informed consent for participation in the study.

Statistical analysis

Intra- and inter-reader reliability of the new scoring systems were assessed using weighted Light's kappa coefficients. This statistical measure takes into account the magnitude of the disagreement between the categories being rated by the different raters, assigning different weights to the disagreements based on the magnitude of the discrepancy between the categories. Kappa values were interpreted according to Landis and Koch[19] as follows: kappa values of 0 - 0.20 were considered as slight, 0.21 - 0.40 as fair, 0.41 - 0.60 as moderate, 0.61 - 0.80 as substantial, and 0.81 - 1.00 as almost perfect agreement.

The number of images to be used in the static images exercise was determined to achieve an expected kappa value of 0.8, with an expected confidence interval lower bound of 0.75. For the patients-based exercise, the reliability was assessed both for specific structures and at patient level (considering all the structures). To reach an expected kappa value of 0.75, with an expected confidence interval lower bound of 0.6, the total number of necessary images for each structure

was 16. The minimum needed number of raters for 16 images was five, but we chose to involve eight raters to be more conservative[20]. Statistical analyses were performed using R statistical software (Foundation for Statistical Computing, Vienna, Austria).

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the manuscript.

Results

Delphi exercise

Based on the results of the preliminary survey, the Delphi questionnaire was designed and consisted of 32 statements which were categorized into 4 main groups. (1) The 1st group focused on the elementary US lesions of CPPD to be applied in the scoring system. (2) The 2nd group dealt with the grading of CPPD, including the general characteristics of the scoring system and the grading of lesions in each structure. (3) The 3rd group focused on the scoring attribution, including the modality of the scoring and the scanning technique. (4) Finally, the 4th group was about the composition of the scoring system, encompassing the joints to be included, the scoring at joint level, and the final scoring at patient level. Three Delphi rounds were needed to reach agreement on all items, with 32 out of 41 members of the OMERACT US WG – CPPD subgroup participating in the first round, 26/32 in the second, and 25/26 in the third round. Overall, participants came from 16 countries, including 10 from Europe, 4 from America, 1 from Africa, and 1 from Oceania. Fourteen statements were approved in the first round, comprising 1 statement each in the 1st and 2nd groups, 4 in the 3rd group, and 8 in the 4th group. No statements were approved in the second round, and 3 statements achieved consensus in the third round, including 1 about the grading of CPPD (1st group) and 2 about the scoring attribution (4th group). The agreed final statements with the percentage of agreement are presented in **Table 1**, all the statements of the Delphi exercise are presented in Supplementary Table S1 in the Appendix page 3.

For the elementary US lesion of CPPD (1st group), the experts decided to apply the OMERACT definitions for US detection of CPPD at the level of HC cartilage and fibrocartilage[12,13] for the

scoring system. Regarding the grading of CPPD (2nd group), the participants agreed upon a semiquantitative scoring system on 4 grades (0 – 3). Grade 0 corresponds to: no findings consistent with CPPD; grade 1: ≤3 single spots or 1 small deposit; grade 2: > 3 single spots or >1 small deposit or ≥1 larger deposit occupying ≤ 50% of the structure under examination in the reference image (i.e., the scanning view with the highest grade of depositions); grade 3: deposits that occupy more than 50% of the structure under examination in the reference image. Additionally, in terms of scoring attribution (3rd group), the evaluation of CPP deposits should be performed on a large portion of the structure under examination, according to the US acoustic window, evaluating the entire structure without lifting the probe (for example entire HC of the femur). The scoring of CPP deposition should be assessed on the scanning view with the highest grade of depositions (reference image). Each structure should be evaluated using a multiplanar approach, scanning at least in two perpendicular planes. Regarding the composition of the scoring system (4th group), the experts agreed that the structures included in the final score were knees, with menisci and HC and the TFCC of the wrists. Each anatomical structure should be scored separately, and then the scores should be summed to define the joint score. The sum of the scores from the assessed joints was the total score at patient level. The final OMERACT US in CPPD scoring system with the definitions and corresponding images are presented in **Figure 1**, and the full scoring template with technical instructions may be found at the supplemental material (Supplementary Figure S1 in the Appendix page 2).

Reliability exercises

For the web-based reliability exercise, six members of the group sent 280 anonymized US images of CPPD patients of the included structures. The organizers (G.F. and S.S.) assessed the images for quality and created a dataset of 120 images on the REDCap platform, representing possibly equally all grades of scoring for each included structure. Thirty-three out of 41 members of the OMERACT US WG - CPPD subgroup completed both rounds of the web-based reliability exercise. The proposed scoring system showed an almost perfect overall reliability for assessing CPPD extent, with a kappa of 0.90 (95% CI 0.79 – 1) for intra-reader reliability, and 0.84 (95% CI 0.79 – 0.88) for inter-reader reliability. All the assessed sites demonstrated an almost perfect intra and inter-reader reliability. The highest agreement was found in HC and menisci, followed by TFCC (HC 0.93 [95% CI 0.78 – 1]

and 0.91 [95% CI 0-85 – 0.94], menisci 0.91 [95% CI 0.79 – 1] and 0.82 [95% CI 0.69 – 0.90], TFCC 0.88 [95% CI 0.71 -1] and 0.81 [95% CI 0.75 – 0.86] respectively) (**Table 2**).

For the patient-based reliability exercise, 8 crystal-proven subjects were enrolled, 50% were female, 100% were white, they have a mean age of 75.3 (±7.9) years and a mean disease duration of 22.5 months (±12.8) from diagnosis (Supplementary Table S2 in the Appendix page 5). On live scanning, the overall intra- and inter-reader reliability of the scoring was substantial, with a kappa of 0.72 (95% CI 0.47-0.96) and 0.66 (95% CI 0.61-0.71), respectively. The most reliable site was HC of the knee, with almost perfect intra- and inter-reader agreement (kappa of 0.87 [95% CI 0.68-1.0] and 0.81 [95% CI 0.61-0.93], respectively), followed by menisci with substantial intra- and inter-reader agreement (kappa of 0.75 [95% CI 0.43-1.0] and 0.70 [95% CI 0.51-0.84], respectively), while the TFCC showed the lowest kappa values, with fair intra- and inter-reader agreement (kappa of 0.35 [95% CI -0.25-0.94] and 0.34 [95% CI 0.13-0.55], respectively). Considering the transverse scan on the menisci in comparison to the longitudinal scan, we found no difference in the agreement, in particular the kappa values for the intra-reader reliability were 0.74 (95% CI 0.45-1.0) and 0.73 (95% CI 0.38-1.0), respectively, for the transverse and longitudinal scans, and the inter-reader kappa values were 0.73 (95% CI 0.59-0.84) for the transverse scan and 0.72 (95% CI 0.49-0.84) for the longitudinal scan on menisci. The results of the reliability of the scoring system on patients are summarized in **Table 2**. The prevalence of the different grades of the scoring system as scored by the readers on the CPPD patients are reported in **Figure 2**.

Regarding the feasibility of the scoring in patients, all examiners completed the four joints examination in less than the provided 12 minutes on all occasions, with a range of scanning times from 6 to 11 minutes max, depending on the difficulty to score.

Discussion

Imaging techniques have greatly improved the understanding of CPPD, but there is a lack of validated instruments for monitoring patients over time. Conventional radiography, is not suitable for evaluating the extent of the disease due to its two-dimensional properties and low sensitivity[21]. As for emerging sophisticated techniques, such as dual energy computed tomography (DECT) or multi-energy/spectral photon-counting CT (SPCCT), even if they appear promising, there are still limited data on their diagnostic performance and their ability to

discriminate between different calcium crystals[22–24]. Further, while US can be performed directly by a rheumatologist during clinical evaluation, other imaging techniques pose a referral problem. Finally, these techniques require radiation exposure and increase direct costs for the management of patients.

This is the first attempt to develop a consensus-based US scoring system for grading CPPD at joint and patient level. According to the OMERACT US framework, the development of the US scoring system was based on data deriving from a SLR[16] and on experts' opinion, since there are no validated scoring systems available in the literature.

For the definition of the scoring system, the experts opted for a four-grades semi-quantitative scoring system from 0 to 3. This decision likely reflects the experts' confidence in four-degree scoring systems, which are also used in other rheumatic diseases[25], and that are generally thought to represent the best compromise between reliability and sensitivity to change.

One of the major challenges encountered in developing the scoring system was to define the different grades. The group agreed on a semi-quantitative definition based on percentages, which may strengthen the reliability and feasibility of the scoring as compared to more subjective definitions. This may appear to be in contrast with what was decided for the OMERACT US gout scoring system, where the definitions of the four grades are based on a qualitative assessment of the item (absent, possible, definite but minimal, definite and severe)[26]. However, in the case of CPPD, crystals develop and grow within an anatomical structure with definite volume (HC, meniscus, TFCC), while tophi and urate crystal deposition on HC (double contour sign) develop and grow within the joint space and may reach considerable volume/extent that cannot be expressed in a percentage of occupation of the joint volume.

Regarding the scoring attribution, the structure under examination should be evaluated entirely, providing as a final score the grade of the frame with the highest extent of the entire structure (reference frame). Although this approach may lead to a loss of information and to a "ceiling effect", it yielded good reproducibility and probably may increase the feasibility of the scoring. A scoring on the entire structure would be challenging, obliging the examiner to make a sort of "mean" of all the frames examined and probably leading to a lower reliability.

The group decided to include in the scoring system only the knees (menisci and femoral HC) and the TFCC of the wrists, as these structures are the most frequently involved joints in CPPD[16], and are

the only sites that have demonstrated high reliability on US[13,27]. In comparison to the 2013 scoring system by Filippou *et al*. that evaluated 26 structures[15], the experts selected a more feasible approach, appropriate also for use in daily practice, including only 8 structures. Tendons were excluded from the final scoring, due to low reliability achieved in previous studies[13].Although this might decrease the sensitivity to change of the scoring, the perception is that a feasible approach is fundamental for clinicians and patients. The instrument demonstrated to be rapid and highly acceptable by the patients, requiring about 10 minutes.

The validation of the scoring system revealed almost perfect intra- and inter-reader reliability on static images, as well as substantial intra- and inter-reader reliability on patients, indicating that the score is clear, feasible and easily applicable. The highest values of intra- and inter-reader reliability were found for HC, followed by menisci. TFCC demonstrated to be the most challenging structure to score on patients, resulting in lower kappa values than the web-based exercise. This difference in agreement suggests that the main difficulty was not the application of the definition, but rather the examination of the TFCC by US. In contrast, the meniscus is generally easier to assess and the scanning technique, as well as tips and tricks for correct crystal identification have been published previously[18,28,29]. The TFCC is a complex structure located in the ulnar side of the wrist between the ulna, lunate and triquetrum bones, composed by an articular disc, a meniscus homologue and several ligaments. US scanning is challenging as the orientation of the different structures that compose the TFCC cover different anatomical planes and, depending on the position of the wrist and of the probe, ligamentous structures present different degrees of echogenicity. Further the meniscus homologue cannot be clearly identified and the echogenicity of the TFCC may change over life time due to degeneration of the structures [30]. At US and in normal conditions the TFCC appears as an inhomogeneous mid-grey (isoechoic) or dark (hypoechoic) area depending on the position of the probe, wrist and tightness of the ligaments at that given point. However, the ligaments that cross the structure may also appear hyperechogenic, and the presence of abnormalities such as fibrocartilage degeneration or areas of fibrosis could mimic CPPD and lead to misinterpretations. CPP deposits appear bright with an echogenicity similar to the bony cortex and can be identified independently of the angle on insonation[29]. In **Figure 3** we provide some pictorial material on the scanning of the TFCC that may help to identify correctly CPPD and thus improve reliability of the scoring at this specific area. It should be noted anyway that the reliability of the overall scoring system at patient level was not affected by the low reliability founded for TFCC.

Additionally, the reproducibility of US assessment is influenced by a complex interaction between US machines, patients, and operators. To minimize the impact of the variability in US machines we used similar high-end US machines equipped with the same probes and settings. Although this may be seen as a limitation of the study, it allowed to eliminate the impact of the machine quality on reliability. Further studies with different machines will allow to explore also the impact of the machine quality on the level of agreement.

In 2018, the Gout, Hyperuricemia and Crystal Associated Disease Network (G-CAN) identified a list of unmet needs in CPPD[5], which included the need for further research on imaging and the improvement of internationally accepted and standardized imaging protocols. US, with the development of standardized definitions and the here presented consensus-based scoring system by the OMERACT US WG - CPPD subgroup, is currently the most validated imaging technique in CPPD. This novel scoring system could be a new tool in the research path on CPPD. An US follow-up could yield new evidence about key questions in CPPD, such as the natural history of the disease, whether and how CPP crystals deposition might change over time, for example following therapy or after an acute attack. Further, a major unmet need, namely the relationship between deposition extent and clinical manifestations or severity of osteoarthritis could be investigated. Finally, one of the most important research priorities in CPPD disease is the development of targeted therapies, capable to inhibit or reverse CPP crystal formation and deposition. The presented scoring system, could be useful in the assessment of changes of deposition after an intervention[31].

The major advantage of this new US scoring system is the development under the OMERACT umbrella. This guarantees the internationality of the faculty, the expertise and mainly a rigorous methodology that has been developed and refined over years of exceptional work and results. The Delphi approach used for the development of the scoring system guarantees the anonymity of all participants, ensuring that each participant's input carries equal weight in the final decision[32,33]. On the other hand, this kind of endeavors are affected by some limitations. The high expertise of the faculty and the quality of machines could apparently make the results non generalizable to the real world. However, the CPPD scoring system is based on grey-scale findings that are less prone to inter-machine variability as compared to Doppler, and with some basic training, the definitions have demonstrated to be reliable also in an external context[34]. Future studies will better assess the feasibility and reliability of this scoring system in a real-world setting.

In conclusion, this study constitutes the first attempt to generate a validated consensus-based semi-quantitative US scoring system for the extent of CPPD. This scoring system is feasible and may ensure valid and comparable results in future clinical trials. It could provide relevant insights into CPPD pathogenesis and natural history, as well as about the potential role of CPPD on osteoarthritis and joint damage and could be useful in assessing the effectiveness of potential therapies. Future studies will assess the sensitivity to change of the scoring system in independent cohorts and evaluate its reliability and feasibility in a real-world setting for external validation. With this step, US has emerged as the better validated technique for CPPD identification and monitoring.

Acknowledgements: The authors would like to thank Gilberto Cincinelli, Raffaele Di-Taranto, Andrea Locatelli, Alessandro Lucia, Andrea Rindone, (Department of Clinical Sciences and Community Health, Università degli Studi di Milano, Milan, Italy), Marwin Gutierrez (Division of Rheumatology, Instituto Nacional de Rehabilitacion, Mexico City, Mexico), Mihaela Micu (Rheumatology Division, 2nd Rehabilitation Department, Rehabilitation Clinical Hospital, Cluj-Napoca, Romania), Giovanni Pacini (Division of Rheumatology, Department of Experimental and Clinical Medicine, University of Florence, Italy), Valentina Picerno, Carmela Toscano (Istituto Reumatologico Lucano (I.Re.L), Dipartimento Regionale di Reumatologia, AOR San Carlo di Potenza, Potenza, Italy), Garifallia Sakellariou (Department of Internal Medicine and Therapeutics, Università di Pavia, and Istituti Clinici Scientifici Maugeri IRCCS, Pavia, Italy), Carlo-Alberto Scirè (Società Italiana di Reumatologia, Epidemiology Research Unit, and School of Medicine, University of Milano-Bicocca, Milan, Italy), Violeta Vlad (Sf. Maria Hospital, Rheumatology Department, Bucharest, Romania), for the contributions in the work, and greatly acknowledge Samsung Electronics Italia S.p.A for providing the US equipment for this study.

This paper was supported and funded by the Italian Ministry of Health - "Ricerca Corrente".

Contributors: GF conceived the study, oversaw study design and results interpretation. SS contributed to study design, performed data analysis, results interpretations and drafted the first version of the manuscript. AZ and AD aid with data analysis and results interpretations. All authors were involved in data acquisitions, provided comments on the manuscript and approved the final version of the manuscript. All authors accept full responsibility for the work, had access to the data, and controlled the decision to publish.

Declaration of interests: all authors declare no competing interest.

Patient and public involvement: Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Ethics approval: The study was approved by the institutional ethics committee of Luigi Sacco University Hospital, Milan, Italy (Registro Sperimentazioni n. 2022/ST/145). Participants gave informed consent to participate in the study before taking part.

Data sharing statement: Data are available upon reasonable request. Data of the paper are deidentified participant data and are available upon request to the correspondent author.

References

1 Zhang W, Doherty M, Bardin T, *et al.* European League Against Rheumatism recommendations for calcium pyrophosphate deposition. Part I: terminology and diagnosis. *Ann Rheum Dis* 2011;**70**:563–70. doi:10.1136/ard.2010.139105

2 Ciancio G, Bortoluzzi A, Govoni M. Epidemiology of gout and chondrocalcinosis. *Reumatismo* 2011;**63**:207–20.

3 Abhishek A. Calcium pyrophosphate deposition disease: a review of epidemiologic findings. *Current Opinion in Rheumatology* 2016;**28**:133–9. doi:10.1097/BOR.0000000000000246

4 Mccarty DJ, Kohn NN, Faires JS. The Significance of Calcium Phosphate Crystals in the Synovial Fluid of Arthritic Patients: The “Pseudogout Syndrome.” *Ann Intern Med* 1962;**56**:711–37. doi:10.7326/0003-4819-56-5-711

5 Abhishek A, Neogi T, Choi H, *et al.* Review: Unmet Needs and the Path Forward in Joint Disease Associated With Calcium Pyrophosphate Crystal Deposition. *Arthritis Rheumatol* 2018;**70**:1182–91. doi:10.1002/art.40517

6 Coari G, Iagnocco A, Zoppini A. Chondrocalcinosis: Sonographic study of the knee. *CLIN RHEUMATOL* 1995;**14**:511–4. doi:10.1007/BF02208146

7 Filippou G, Frediani B, Gallo A, *et al.* A “new” technique for the diagnosis of chondrocalcinosis of the knee: Sensitivity and specificity of high-frequency ultrasonography [5]. *Ann Rheum Dis* 2007;**66**:1126–8. doi:10.1136/ard.2007.069344

8 Kellner H, Zoller W, Herzer P. [Ultrasound findings in chondrocalcinosis]. *Z Rheumatol* 1990;**49**:147–50.

9 Filippou G, Adinolfi A, Cimmino MA, *et al.* Diagnostic accuracy of ultrasound, conventional radiography and synovial fluid analysis in the diagnosis of calcium pyrophosphate dihydrate crystal deposition disease. *Clin Exp Rheumatol* 2016;**34**:254–60.

10 Cipolletta E, Filippou G, Scirè CA, *et al.* The diagnostic value of conventional radiography and musculoskeletal ultrasonography in calcium pyrophosphate deposition disease: a systematic literature review and meta-analysis. *Osteoarthritis and Cartilage* 2021;**29**:619–32. doi:10.1016/j.joca.2021.01.007

11 Terslev L, Naredo E, Keen HI, *et al.* The OMERACT Stepwise Approach to Select and Develop Imaging Outcome Measurement Instruments: The Musculoskeletal Ultrasound Example. *J Rheumatol* 2019;**46**:1394–400. doi:10.3899/jrheum.181158

12 Filippou G, Scirè CA, Damjanov N, *et al.* Definition and Reliability Assessment of Elementary Ultrasonographic Findings in Calcium Pyrophosphate Deposition Disease: A Study by the OMERACT Calcium Pyrophosphate Deposition Disease Ultrasound Subtask Force. *J Rheumatol* 2017;**44**:1744–9. doi:10.3899/jrheum.161057

13 Filippou G, Scirè CA, Adinolfi A, *et al.* Identification of calcium pyrophosphate deposition disease (CPPD) by ultrasound: reliability of the OMERACT definitions in an extended set of

joints—an international multiobserver study by the OMERACT Calcium Pyrophosphate Deposition Disease Ultrasound Subtask Force. *Annals of the Rheumatic Diseases* 2018;**77**:1194–9. doi:10.1136/annrheumdis-2017-212542

14 Filippou G, Scanu A, Adinolfi A, *et al.* Criterion validity of ultrasound in the identification of calcium pyrophosphate crystal deposits at the knee: an OMERACT ultrasound study. *Ann Rheum Dis* 2020;:annrheumdis-2020-217998. doi:10.1136/annrheumdis-2020-217998

15 Filippou G, Filippucci E, Tardella M, *et al.* Extent and distribution of CPP deposits in patients affected by calcium pyrophosphate dihydrate deposition disease: An ultrasonographic study. *Ann Rheum Dis* 2013;**72**:1836–9. doi:10.1136/annrheumdis-2012-202748

16 Adinolfi A, Sirotti S, Sakellariou G, *et al.* Which are the most frequently involved peripheral joints in calcium pyrophosphate crystal deposition at imaging? A systematic literature review and meta-analysis by the OMERACT ultrasound – CPPD subgroup. *Frontiers in Medicine* 2023;**10**.https://www.frontiersin.org/articles/10.3389/fmed.2023.1131362 (accessed 11 Mar 2023).

17 Möller I, Janta I, Backhaus M, *et al.* The 2017 EULAR standardised procedures for ultrasound imaging in rheumatology. *Annals of the Rheumatic Diseases* 2017;:annrheumdis-2017-211585. doi:10.1136/annrheumdis-2017-211585

18 Filippou G, Picerno V, Adinolfi A, *et al.* Change perspective to increase diagnostic accuracy of ultrasonography in calcium pyrophosphate dihydrate deposition disease! A new approach: the axial scan of the meniscus. *Reumatismo* 2015;**66**:318–21.

19 Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. *Biometrics* 1977;**33**:159. doi:10.2307/2529310

20 Rotondi MA, Donner A. A confidence interval approach to sample size estimation for interobserver agreement studies with multiple raters and outcomes. *J Clin Epidemiol* 2012;**65**:778–84. doi:10.1016/j.jclinepi.2011.10.019

21 Sirotti S, Becce F, Sconfienza LM, *et al.* Reliability and diagnostic accuracy of radiography for the diagnosis of calcium pyrophosphate deposition: performance of the novel definitions developed by an international multidisciplinary working group. *Arthritis & Rheumatology*;**n/a**. doi:10.1002/art.42368

22 Budzik J-F, Marzin C, Legrand J, *et al.* Can Dual-Energy Computed Tomography Be Used to Identify Early Calcium Crystal Deposition in the Knees of Patients With Calcium Pyrophosphate Deposition? *Arthritis Rheumatol* 2021;**73**:687–92. doi:10.1002/art.41569

23 Stamp LK, Anderson NG, Becce F, *et al.* Clinical Utility of Multi-Energy Spectral Photon-Counting Computed Tomography in Crystal Arthritis. *Arthritis & Rheumatology* 2019;**71**:1158–62. doi:10.1002/art.40848

24 Filippou G, Pascart T, Iagnocco A. Utility of Ultrasound and Dual Energy CT in Crystal Disease Diagnosis and Management. *Curr Rheumatol Rep* 2020;**22**:15. doi:10.1007/s11926-020-0890-1

25 Bruyn GA, Iagnocco A, Naredo E, *et al.* OMERACT Definitions for Ultrasonographic Pathology and Elementary Lesions Of Rheumatic Disorders Fifteen Years On. *The Journal of Rheumatology* 2019;:jrheum.181095. doi:10.3899/jrheum.181095

26 Christiansen SN, Filippou G, Scirè CA, *et al.* Consensus-based semi-quantitative ultrasound scoring system for gout lesions: Results of an OMERACT Delphi process and web-reliability exercise. *Seminars in Arthritis and Rheumatism* 2021;:S0049017221000056. doi:10.1016/j.semarthrit.2020.11.011

27 Filippou G, Scanu A, Adinolfi A, *et al.* Criterion validity of ultrasound in the identification of calcium pyrophosphate crystal deposits at the knee: an OMERACT ultrasound study. *Ann Rheum Dis* Published Online First: 2020. doi:10.1136/annrheumdis-2020-217998

28 Filippou G, Adinolfi A, Bozios P, *et al.* Do not hallow until you are out of the wood! Ultrasonographic detection of cpp crystal deposits in menisci: Facts and pitfalls. *Sci World J* 2013;**2013**. doi:10.1155/2013/181826

29 Filippou G, Sirotti S. How can ultrasonography help in the management of CPPD? From diagnosis to clinical subset identification. *Current Opinion in Rheumatology* 2023;**35**:185–93. doi:10.1097/BOR.0000000000000939

30 Wu W-T, Chang K-V, Mezian K, *et al.* Ulnar Wrist Pain Revisited: Ultrasound Diagnosis and Guided Injection for Triangular Fibrocartilage Complex Injuries. *JCM* 2019;**8**:1540. doi:10.3390/jcm8101540

31 Cai K, Tedeschi SK. Review: Outcome measures in calcium pyrophosphate deposition. *Best Practice & Research Clinical Rheumatology* 2021;**35**:101724. doi:10.1016/j.berh.2021.101724

32 Hsu C-C, Sandford BA. The Delphi Technique: Making Sense of Consensus. doi:10.7275/PDZ9-TH90

33 Raine R, Sanderson C, Black N. Developing clinical guidelines: a challenge to current methods. *BMJ* 2005;**331**:631–3.

34 Lee K-A, Lee S-H, Kim H-R. Diagnostic value of ultrasound in calcium pyrophosphate deposition disease of the knee joint. *Osteoarthritis and Cartilage* 2019;**27**:781–7. doi:10.1016/j.joca.2018.11.013