Review

International Journal of Clinical Rheumatology

Biomarkers for knee osteoarthritis: new technologies, new paradigms

There is a paucity of biomarkers in knee osteoarthritis (OA) to inform clinical decision making, evaluate treatments, enable early detection and identify people who are most likely to progress to severe OA. The absence of biomarkers places considerable limitations on the design of research studies, and is a barrier toward applying the principles of stratified medicine to knee OA. Here we describe key principles and processes of biomarker development and focus on two promising areas that draw upon technologies that have only relatively recently been developed for quantitative applications in clinical research, namely 3D MRI and acoustic emission. While still at an early stage, results to date show promising potential to open up interesting new paradigms in this field.

Keywords: 3D MRI • acoustic emission • biomarker • knee osteoarthritis

Lucy Spain[1], Bashar Rajoub[2], Daniela K Schlüter[3], John Waterton[4], Mike Bowes[5], Lik-Kwan Shark[6], Peter Diggle[7] & John Goodacre[*,8]

[1]Clinical Research Associate, Lancaster Health Hub, Faculty of Health & Medicine, Lancaster University, UK
[2]Research Associate, Applied Digital Signal & Image Processing Research Center (ADSIP), School of Computing, Engineering & Physical Sciences, University of Central Lancashire, UK
[3]Senior Research Associate, CHICAS, Lancaster Medical School, Lancaster University, UK
[4]Professor of Translational Imaging, Biomedical Imaging Institute, Manchester Academic Health Sciences Center, University of Manchester, UK
[5]CEO, Imorphics Ltd, Manchester Science Park, Manchester, UK
[6]Head of Applied Digital Signal & Image Processing Research Center (ADSIP), School of Computing, Engineering & Physical Sciences, University of Central Lancashire, UK
[7]Head of CHICAS, Lancaster Medical School, Lancaster University, UK
[8]Professor of Musculoskeletal Science, Lancaster Health Hub, Faculty of Health & Medicine, Lancaster University, Lancaster, LA1 4YG, UK
*Author for correspondence: j.goodacre@lancaster.ac.uk

Developing new biomarkers: general principles & approaches

Knee osteoarthritis (OA) is a common but heterogeneous condition. The diagnosis of knee OA continues to rely heavily on x-radiology, and is based upon a combination of characteristic structural features and pain symptoms [1]. However, x-ray features correlate relatively poorly with pain symptoms [2,3] and are of limited value in the early stages. In contrast with many other conditions, such as rheumatoid arthritis, diabetes and cancer, there is a paucity of biomarkers in knee OA to inform clinical decision making and to enable evaluation of treatments and other interventions. Furthermore, there are no biomarkers either to enable early detection of the condition, or to identify people who are most likely to progress to severe OA. The absence of biomarkers also limits the range of available research approaches for gaining a better understanding of the underlying biology of knee OA, and places considerable limitations on options for designing studies to evaluate new treatments, such as cartilage regeneration. Also, given the recognized clinical and biological heterogeneity of knee OA, there is an urgent need for biomarkers to enable the principles of stratified medicine to be explored and applied for this condition.

Biomarkers are characteristics that are 'objectively measured and evaluated as indicators of normal biological processes, pathogenic processes or pharmacologic responses to a therapeutic intervention' [4,5]. This definition allows biomarkers to be either numerical (e.g., joint space width/mm; volume of medial tibial cartilage/ml) or categorical (e.g., Kellgren-Lawrence (KL) grade; MOAKS synovitis score [6]), so long as they are objectively measured. In principle, a perfect biomarker correctly predicts clinical outcome [7]. Since no biomarker is perfectly valid, investigation and development of a new candidate biomarker involves a wide range of activities to ensure that the uncertainty, risk and cost in making research or clinical decisions reliant on the biomarker can be managed. Such activities involve:

- Technical validation, based on the concept that measurements made anywhere in the world should be identical or acceptably similar. These activities can be subdivided

Future Medicine part of fsg

into repeatability, reproducibility and availability. Repeatability is the idea that measurements should be similar when made on the same person, by the same operator, using the same equipment and software, in the same setting, over a short period of time. Reproducibility is the idea that measurements should be similar when made on equivalent subjects, by different operators, using different equipment, in different settings, at different times. Availability is the idea that there are no legal, ethical, regulatory or commercial barriers preventing measurement in particular settings or jurisdictions. Technical validation does not, of itself, provide any evidence that the biomarker is useful. However, it is a prerequisite for large multicenter studies or meta-analyses.

- Biological and clinical validation, based upon the concepts that the biomarker should faithfully represent underlying biology, and accurately forecast clinical outcome. This has been described as a 'graded evidentiary process' dependent on the intended application [8]. Biological validation does not, of itself, assume that the biomarker can be robustly measured in multiple centers.

Much of the academic and regulatory literature on biomarker validation aligns with 'biospecimen' biomarkers derived from patients' tissues or biofluids, where the biomarker is a specified molecular entity whose measurement is essentially an exercise in analytical biochemistry. However, there is considerably less literature regarding validation of 'biosignal' biomarkers from, for example, imaging, acoustic emission or electrophysiology. Biospecimen biomarkers are typically measured using a dedicated *in vitro* diagnostic device remote from the patient, where the device's performance can be optimized on historical or biobanked samples. For biosignal biomarkers on the other hand, technical quality depends mainly on how the measurement is performed while the patient is physically present and coupled to the *in vivo* diagnostic device. Biosignal repeatability may be relatively easily achieved, allowing small studies by a single investigator. However, multicenter reproducibility is often challenging, because of the need for training and standardization of the biosignal device and its use in each setting.

For biospecimen biomarkers, both in guidelines and in actual practice, technical (assay) validation can mostly be achieved at a fairly early stage, possibly with some early clinical validation from biobanked samples where the patient outcome is known. The biomarker already has great credibility before a single new patient is recruited. For biosignal biomarkers, on the other hand, a stepwise approach [9], where small increments of technical and biological–clinical validation are addressed in parallel, is more appropriate. Early studies address repeatability in single centers, or in a few centers using identical equipment. At the same time, biological and clinical validity may be approached tentatively using the Bradford Hill [9–11] criteria, for example with small cross-sectional or interventional studies. Only later, after extensive efforts to establish that the biomarker measures are the same in different centers, using different equipment in different jurisdictions, can definitive outcome studies be performed.

Developing new biomarkers: key statistical issues

As described above, the process for developing and validating new biomarkers is complex and multistaged. Broadly speaking, the key statistical issues that arise in any study aimed at developing a novel biomarker for a chronic condition depend crucially upon the stage of biomarker development on which the study is focused. Among these, however, the following are highlighted as being particularly important questions in relation to this issue:

Is the candidate biomarker repeatable & reproducible?

In the early stages of biomarker development, the focus may be on repeatability and reproducibility, in other words, roughly speaking, S/N ratio, where 'noise' can encompass technical variation in the measurement device, short-term, clinically irrelevant biological variation in the patient, and variation between multiple observers.

How does the candidate biomarker compare with other biomarkers?

Once a candidate biomarker has passed the initial test described above, the next step focuses on comparing it in cross-sectional studies with other, more established, biomarkers for the same condition. An important issue is then the level of validity of the current default biomarker as a 'gold standard'. If the validity of the current 'gold standard' biomarker is high (i.e., approaching the Prentice criteria for surrogacy), the candidate biomarker will likely be judged by the extent to which it can (almost) match the current default's predictions but at substantially lower cost or greater convenience. If, however, the validity of the current biomarker is relatively low, the emphasis for the candidate biomarker is more likely to be on improving predictive performance.

Can the candidate biomarker predict clinical outcomes?

In either case, if the candidate biomarker is still in the frame after initial testing, its most severe test is its

ability to predict important clinical outcomes, so as to justify its use either as a surrogate end point in clinical trials, or for use in clinical practice, perhaps to allow an intervention intended to reverse, or at least slow, clinical progression. A good example of a biomarker that has proven value as an early indicator of the need for clinical action is the rate of change in serum creatinine as a biomarker for incipient renal failure [12,13].

Surrogate end points have a chequered history. In the statistical literature, the widely cited 'Prentice criteria' for surrogacy [7] are very difficult to establish in practice. In particular, establishing a correlation between a health outcome and a biomarker is far from sufficient to establish surrogacy [14]. In the medical literature, Psaty *et al.* [15] discuss how treatments that appear beneficial on the basis of a surrogate end point can later prove to have a harmful effect on the relevant clinical end point.

Modern technological developments are increasingly opening up new possibilities for biomarker discovery. Traditionally, the term 'biomarker' would refer to a direct biophysical or biochemical measurement. Nowadays, however, the term can include summary descriptors of an intrinsically high-dimensional object, such as a digital image, a time series or a chemical spectrum, termed 'biosignals'. Our work to explore acoustic emission (AE) as a biomarker for knee OA [16–19] illustrates some of the challenges which are being raised by the ever-broadening scope of technologies being explored as potential biosignals.

As described below, the raw data from an AE time series consist of a record of noise levels recorded from knees throughout a person's sit-stand-sit movement. Low levels of AE are potentially contaminated by background noise. For this reason, the recording equipment stores only AE levels that exceed a prespecified threshold. Each such exceedance is termed a 'hit'. Potential biomarkers include the frequency and positions of hits, or summary statistics of the real-valued time series of AE levels during each hit. Formally, this defines a 'marked point process', in which the points are the onset times for each hit, while the marks are the associated time series of AE levels throughout the corresponding excursions over the threshold. The statistical challenge is to formulate and fit to this complex, very high-dimensional object, a model that effects a dramatic reduction in dimensionality from the data to a set of parameter estimates or summary statistics that capture the essential properties of the data.

Once a set of summary statistics has been identified, the next step is to analyse the components of variation in each. Typically, these can include systematic variation between identifiably different groups, for example people with the condition compared with healthy controls, and a hierarchy of random components of variation: between people in the same group; between repeat series on the same person at different times; technical variation between ostensibly equivalent pieces of equipment and inherent statistical variation in the summary statistics (measurement error). Reproducibility requires that variation between groups and between patients within groups should dominate variation between times within patients, technical variation and inherent statistical variation.

This has direct implications for study design. A minimal requirement is that the design includes replication at each level of the hierarchy so that all of the variance components are identifiable. The more ambitious goal of a formal sample size calculation to guarantee acceptable levels of precision in the estimated components of variation is only feasible if the different phases are conducted as separate studies, with the results of the initial study informing the sample size for the subsequent validation study.

When a candidate biomarker has passed the development phase, it can then be tested in further work to determine whether it can predict, with practically useful accuracy, a clinically relevant end point. To avoid the danger of overfitting, the data used for this clinical validation phase must be separate from the data used in the development phase. This can be achieved either by a formal separation into two independent studies, or by dividing the data into two subsets. For the same reason, direct comparison of a candidate biomarker with the clinically relevant end point should be reserved for the clinical validation phase. In our studies, candidate biomarkers are being developed based on properties of the AE signals that can discriminate between painful and pain-free knees, and between other current markers of severity. Clinical validation against disease progression will require longer term follow-up studies.

Developing new biomarkers for knee OA

In knee OA, the current unmet need is for biomarkers which can improve our forecast of clinical outcomes. For example, can we predict who will respond well to NSAIDs, and who will suffer worsening pain and disability leading to early knee replacement? Can we predict which people will benefit most from knee replacement? Can we identify specific cohorts (stratified medicine) who will benefit from interventions such as insoles or high tibial osteotomy, or from investigational new drugs?

Imaging biomarkers based on radiographic features, particularly KL scores, have played a central role in OA research, response assessment and predicting clinical outcome for many years. Other imaging modalities are currently emerging, including CT scanning,

arthroscopy, PET, single-photon emission computed tomography/scintigraphy, ultrasound and MRI. Of these, MRI is particularly promising and offers the prospect of using quantitative measures as biomarkers from several tissues and structures involved in the disease process [20,21], and of demonstrating moderately good correlation with symptoms [22–25]. MRI has the important advantage of avoiding the need for ionizing radiation and of being reasonably widely available for use in clinical trials, although it is difficult to use for population- and primary care-based studies. Therefore, there continues to be a need to explore other potential approaches as well.

Given the paucity of robust biomarkers within this field, and the heterogeneity and slow rate of change of the condition, the development and validation of a new biomarker necessarily requires work to be conducted over several years, and to include the use of large-scale longitudinal studies to inform its adoption. Furthermore, depending on the characteristics of the candidate biomarker, it may prove extremely difficult to fully test a candidate biomarker with the necessary stringency to fully satisfy the requisite criteria. Nevertheless, while this is clearly a very challenging field, encouraging progress is being made in terms of investigating new candidates. The two approaches described here demonstrate potential not only as useful additions to the range of available biomarkers, but also to offer new paradigms for measuring and studying knee OA.

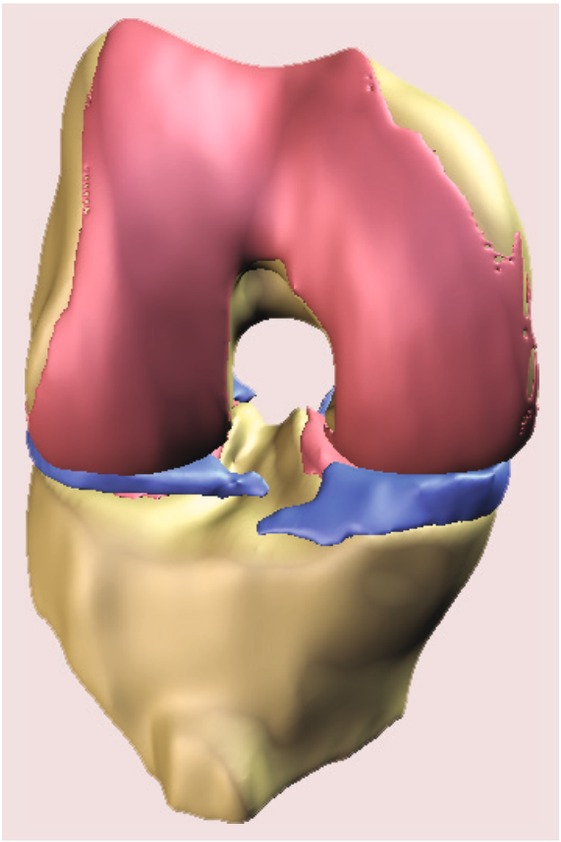

Figure 1. 3D MRI of knee without osteoarthritis.
Representative example of healthy knee with no evidence of knee osteoarthritis (patella not shown). Articular cartilage (pink) covers whole of the articulating surface, there is no evidence of osteophytes around the cartilage plates, and the menisci are full sized.

3D MRI

The increased availability of magnetic resonance scanners in recent years has enabled a number of potential imaging biomarkers for knee OA to emerge. Broadly speaking, there are two approaches to the measurement and interpretation of MRIs of the knee for this purpose:

- Categorical, often semi-quantitative scales such as BLOKS and MOAKS [6].
- Quantitative methods [26].

The primary advantages of semi-quantitative scoring systems are that they consider the whole knee instead of focusing on individual tissues, and exploit the proficiency of expert readers to judge the condition of the whole knee. Their primary disadvantage is that they are time consuming to perform, and demanding for the readers. A further disadvantage is that these scales have multiple subscales, and it is not at all clear whether and how these subscales should be combined.

To date, most of the published work involving quantitative change in the knee has focused on changes in the morphology of articular cartilage in the femorotibial joint. Changes in cartilage thickness and volume have been extensively reported (for example, [27]). These changes have been shown to be responsive, and to predict future OA progression, as well as risk of total knee replacement, and are weakly associated with clinical symptoms such as pain.

Recently, changes in 3D bone shape have been reported in a large group of OA patients, and shown to be more responsive than changes in cartilage [28] thickness and radiographic joint space width. Bone is known to change with the progression of OA, and the tibial condyles have demonstrated an increase in bone area [29,30].

Several other imaging biomarkers have been suggested, and are being actively pursued, including bone marrow lesions, meniscal volume and extrusion, and synovitis. It is not yet known how each of these tissues change with time, and with extent of disease, and how they interact with each other. Each of these imaging biomarkers share common features. OA is a slow-moving disease, and the amount of change which occurs

within a clinical study (typically of 1 year's duration) is very small. For example, cartilage thickness changes by fractions of 1 mm per year on average, when the total thickness is only 3–5 mm. This amount of change represents about a quarter of the width of a single pixel on the screen. Measuring such small changes is a challenge. It requires painstaking work, excellent measurement tools and the use of very large patient cohorts to establish statistical significance.

Interpretation of imaging biomarkers in terms of clinical outcomes is also challenging. The relationship of small changes in tissues does not correlate well with clinical outcomes, such as pain. Interpretation of imaging biomarkers and their relationship to patient outcomes therefore requires a thorough approach and careful analysis. Examples of 3D magnetic resonance images for knee OA are shown in Figures 1 and 2.

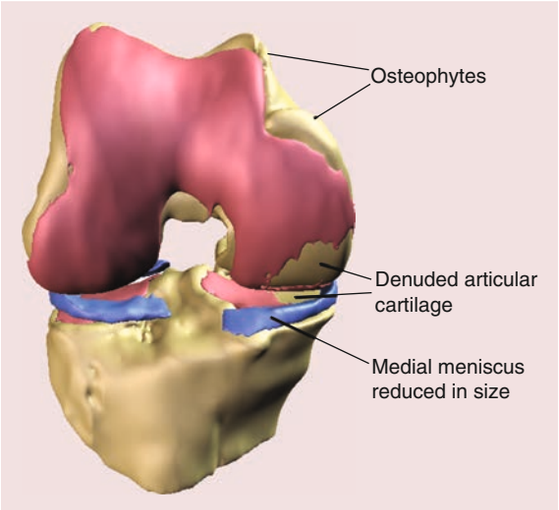

Figure 2. 3D MRI of knee with osteoarthritis.
Representative example of osteoarthritis knee. Articular cartilage is characteristically denuded in the medial femorotibial joint. Osteophytic growth is apparent around most of the femoral cartilage, most prominently on the medial side, and the menisci are smaller in size, possibly as a result of previous meniscal surgery.

Acoustic emission

In recent years there has been increasing interest in translational applications of acoustic science to inform the diagnosis and management of several different health [31–39] and dental conditions [40,41]. In this context, the development of techniques to measure high-frequency AE in knee joints [16–19] offers the potential to develop a biomarker reflecting the integrity of interactions between knee joint components during weight-bearing movement. Since OA affects joint function, markers which assess or reflect the integrity of interactions between joint structures during knee movement offer a strong rationale for biomarker development.

Noninvasive, portable systems have been developed to capture and analyse high-frequency sound emitted during weight-bearing knee movement, involving the translation of leading-edge AE technology into clinical applications. Wide-band AE sensors detect sound waves with frequencies up to 200 kHz emitted from knees during sit-stand-sit movement. Similar principles have been widely utilized for many years in nondestructive testing and condition monitoring of engineering structures for early detection of damage and material defects. By analogy, surfaces which are smooth and well lubricated move quietly against each other, whereas rough, poorly lubricated surfaces move unevenly, producing acoustic signals.

AE sensors are piezoelectric transducers which are attached to knee surfaces, using defined anatomical landmarks, to record short bursts of acoustic energy generated by stress upon, and friction between, joint components during weight-bearing movement.

Data collection involves simultaneous recording of weight distribution, joint angle and acoustic emissions from both knees using a 'Joint Acoustic Analysis System'. AE sensors are positioned anterior to the medial patellar retinaculum and attached to ensure a good connection between source and sensor. Electrogoni-

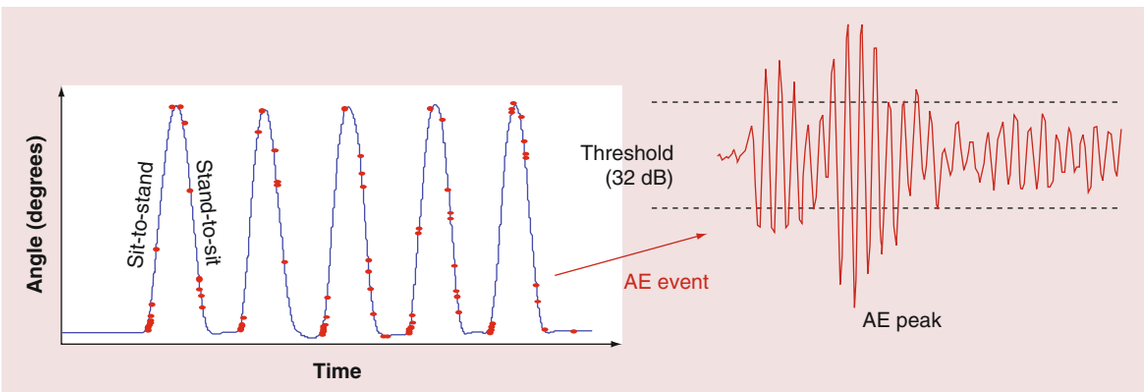

Figure 3. Joint angle-based acoustic emission.

ometers are positioned laterally to each knee, along the plane between the greater trochanter and lateral malleolus. AE, joint angle and ground reaction forces are recorded while participants perform a set of sit-stand-sit movements, starting in a seated position with their back against the chair and knees bent at 90 degrees. Each test involves two sets of five sit-stand-sit movements. This protocol has been found not only to generate meaningful data but also to be convenient, feasible and acceptable for use in clinical settings.

AE signals are characterized by short-duration burst waveforms. AE data acquisition operates in an event-based recording mode to minimize data volume, and a burst signal must have a significant amplitude in order for it to be recorded as an AE event or AE hit. As an example, the AE events recorded over five repeated sit-stand-sit movements from a person are shown as dots superimposed on the joint angle signal (solid curve) in Figure 3, where the AE waveform for one of the AE events is also shown.

Whereas traditional imaging techniques reflect the static anatomical appearance of knee joints at a particular posture, AE captures dynamic interaction among joint components as a result of movements. Each movement action is divided further into its constituent phases, such that each sit-stand-sit cycle is

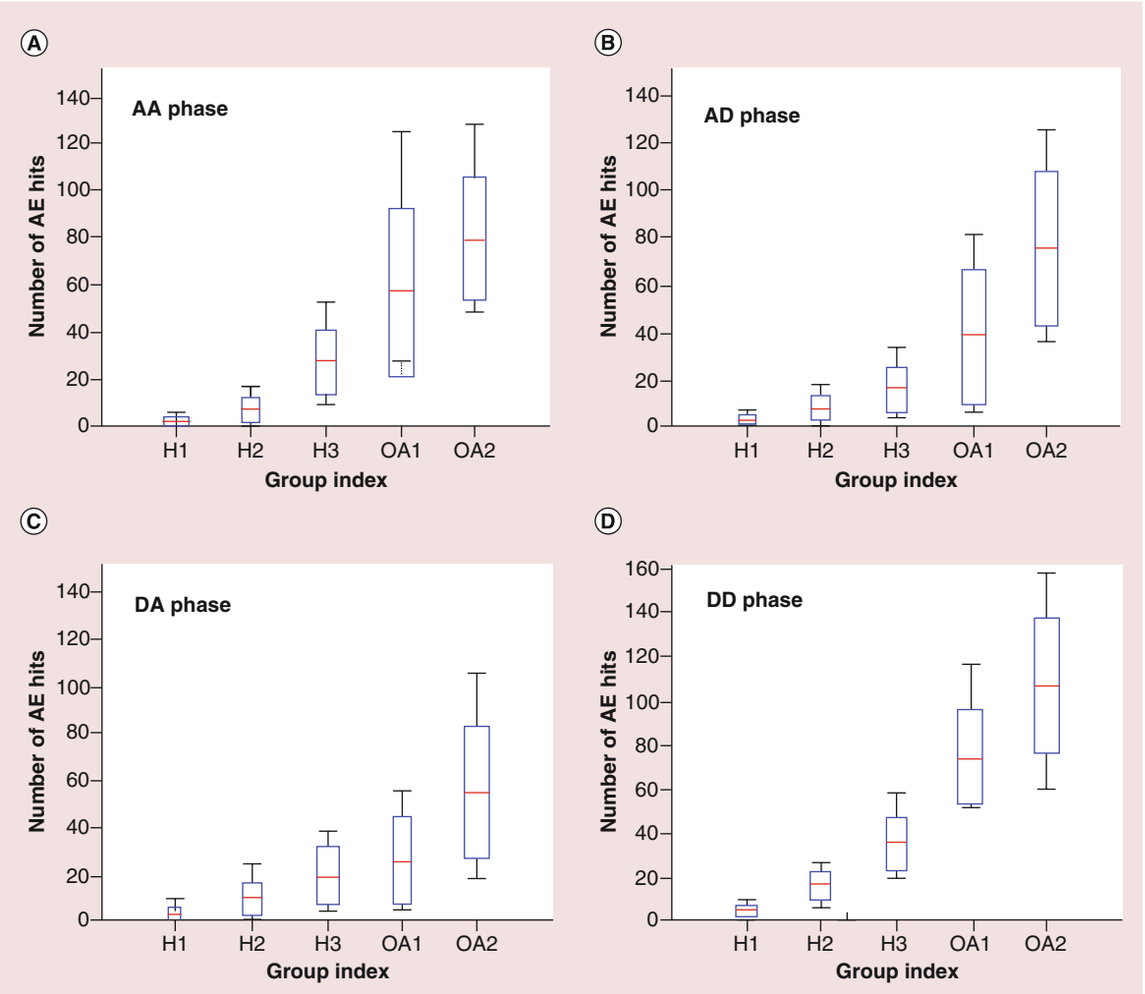

Figure 4. Acoustic emission events in four movement phases in each subgroup.
Group H1: 10 healthy knees from early adulthood aged between 22 and 40 years (mean: 29.00 years; SD: 5.45 years).
Group H2: 11 healthy knees from middle adulthood aged between 42 and 58 years (mean: 50 years; SD: 5.07 years).
Group H3: 13 healthy knees from late adulthood aged over 61 years (mean: 71.27; SD: 6.99 years) (Figure from [18]).
Group OA1: 7 OA knees from middle adulthood aged between 52 and 58 years (mean: 55.00 years; SD: 1.9 years).
Group OA2: 12 OA knees from late adulthood aged over 61 years (mean: 69.5 years; SD: 6.39 years).
AA: Ascending–acceleration phase; AD: Ascending–deceleration phase; DA: Descending–acceleration phase; DD: Descending–deceleration phase denoted by DD; SD: Standard deviation.

divided into the ascending phase (corresponding to the sit-to-stand movement) followed by the descending phase (corresponding to the stand-to-sit movement). Using this two-phase movement model and the average number of AE events per each movement phase, the ascending phase shows the largest difference between young healthy knees and old OA knees. At a finer level, the peak angular velocities derived from the joint angle signal are used to divide each ascending phase and each descending phase into the acceleration phase followed by the deceleration phase, yielding a four-phase movement model for each sit-stand-sit cycle.

Figure 4 shows differences in the number of AE hits in each of four movement phases for five different groups, namely healthy knees from early adulthood (H1), healthy knees from middle adulthood (H2), healthy knees from late adulthood (H3), OA knees from middle adulthood (OA1) and OA knees from late adulthood (OA2). Within this group, the least overlap between the five subgroups occurred in the descending-deceleration (DD) phase of movement.

In order to discover biomarkers, it may be useful to take into account the characteristics of the waveforms which constitute each AE hit. A wide range of waveform features could be used to describe the characteristics of an AE hit. These include amplitude (peak amplitude value of the AE waveform; average signal level [ASL] over waveform duration), energy (sum of absolute or squared amplitude values over waveform duration), frequency (peak frequency of waveform; center and average frequencies over the entire waveform duration; initiation and reverberation frequencies of the waveform before and after the peak) and time (duration of the AE waveform, waveform rise and fall time before and after the peak). Among various AE features investigated to date, AE peak magnitude and ASL have shown good discrimination between subgroups, particularly in the descending-deceleration movement phase.

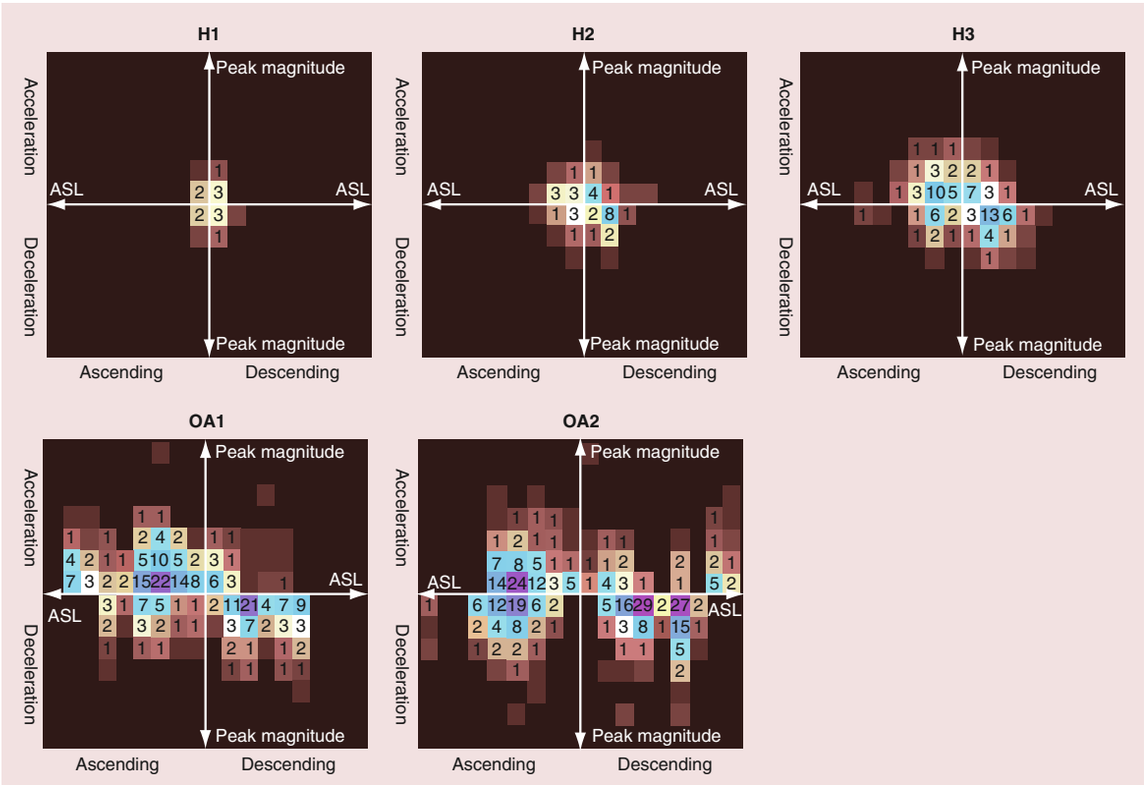

Figure 5. Image-based representation of typical acoustic emission feature profiles. Image-based representation is divided into four quadrants:
Left quadrants show the AE feature profile of the ascending phase.
Right quadrants show the AE feature profile of the descending phase.
Top quadrants show the AE feature profile of the acceleration phase.
Bottom quadrants show the AE feature profile of the deceleration phase.
Within each quadrant, the co-occurrence of AE peak magnitude and ASL values are shown as interval-based 2D histograms with the horizontal axis representing ASL and the vertical axis representing AE peak magnitude.
Figure reproduced with permission from [18].
AE: Acoustic emission; ASL: Average signal level.

Image-based representation has been developed to display the value distributions of AE peak amplitude and ASL features from each knee in each movement phase, thereby allowing visual comparison of different knee joint conditions. Typical AE feature profiles of a knee in each group are shown in Figure 5, in which colors and numerical values indicate the average number of AE hits in each co-occurrence class per movement phase based on two sets of five repeated sit-stand-sit movements. A healthy knee joint in early adulthood (from Group H1) generates a small number of AE hits of lower peak amplitude and ASL values. As age increases, the number of AE hits increase with increasing higher peak amplitude and ASL values (Groups H2 and H3). OA knees generate the highest number of AE hits with a wide range of peak magnitude and ASL values. Furthermore, there are differences in the AE feature profiles between OA knees. AE events with high ASL and low peak amplitude (corresponding to relatively long duration and low magnitude waveforms) occur in the ascending-acceleration movement phase for the OA1 knee and in the descending-acceleration movement phase for the OA2 knee.

Through image-based AE feature profiles, diverse AE occurrences and waveform characteristics are represented in a uniform information format which allows conventional multivariate statistical analysis techniques to be applied to reveal hidden cluster structures without using prior knowledge. One of the most widely used unsupervised methods is principal component analysis (PCA), which is applied to transform a large set of co-occurrence classes in the image-based AE feature profile into a small set of relatively independent variables, known as principal components, in such a way as to highlight the differences and similarities among the measured knee joints.

For example, Figure 6 shows the projection of AE feature profile based on two sets of five repeated sit-stand-sit movements from 53 knee joints to the space defined by the first three principal components, which capture approximately 87.97% of the total variance in the data. By labeling each point projected from each AE feature profile according to its age and knee condition group, five well-separated clusters are seen with a trajectory related to knee age and condition, progressing from Group H1 to Group H3, followed by Group OA1 to OA2. Fur-

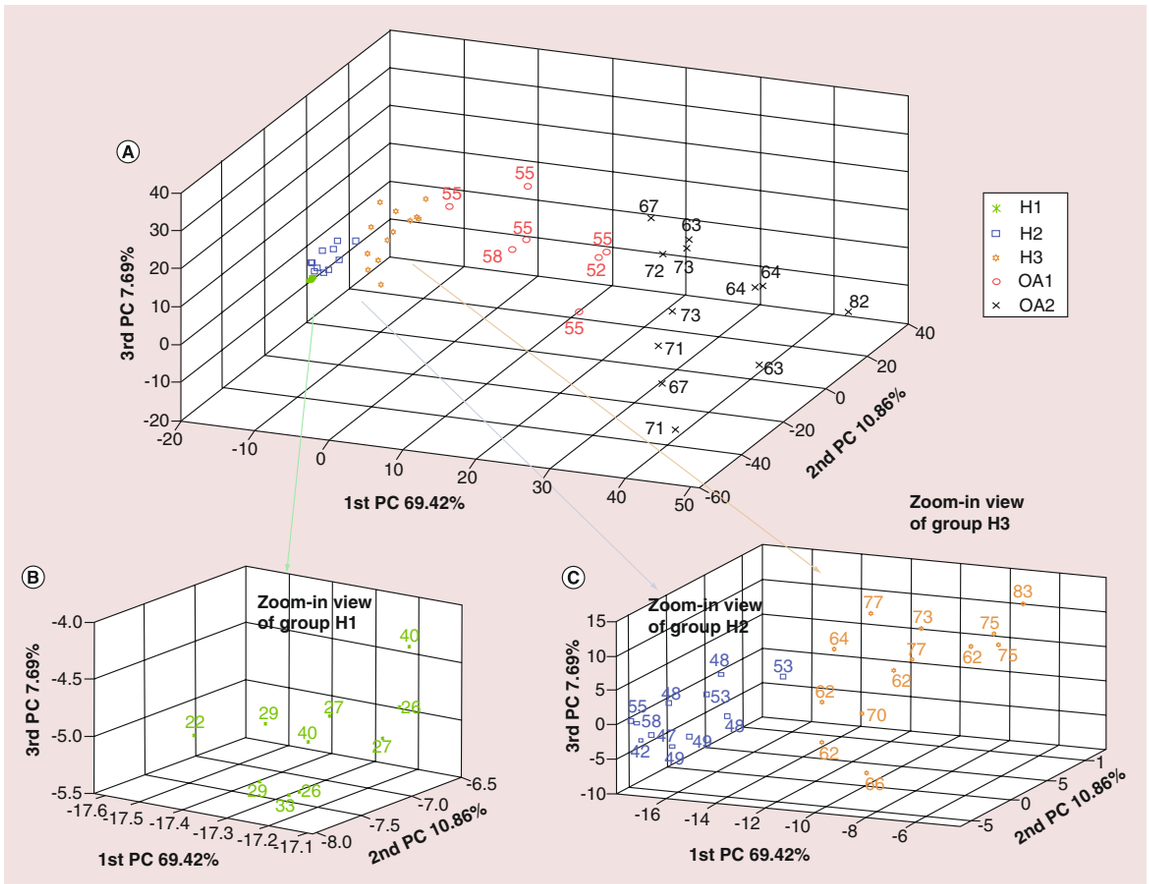

Figure 6. Principal component analysis of acoustic emission feature profiles, in which numbers indicate the age of participants.

thermore, this trajectory shows increasing areas for each cluster, starting from the smallest cluster for Group H1. With increasing age, cluster areas increase with longer distances among the feature profiles for Groups H2 and H3. In changing from healthy to OA, the cluster areas spread even further with longer distances among the projected AE feature profiles of Group OA1 compared with Group H3. Late adulthood OA knees produce the largest cluster area with the widest spread of the projected AE feature profiles. By using the clusters and trajectory in the PCA space as the reference, the condition of a knee can be judged visually based on the projected point of its AE profile along the trajectory and its distance with respect to the center and boundary of the nearest two clusters, and a single quantitative measure can be given based on its PCA coordinates.

These results demonstrate the potential of AE as a biomarker in knee OA. AE offers not only an objective, quantitative measurement tool but also a new paradigm for knee OA assessment and biomarker development. The system is also noninvasive, portable and convenient to use. While a full understanding of the tissue origin of acoustic signals in knee joints awaits further investigation, the findings to date suggest that it may have several potential applications in primary and secondary care settings.

Conclusion & future perspective

Among the increasing array of emerging candidate biomarkers for knee OA, 3D MRI and AE are based on novel principles and technologies, and have demonstrated particularly promising potential for a wide range of applications in research and clinical practice. Currently, studies to assess the precision, sensitivity and inter/intra-user reproducibility of AE and MR imaging compared with radiographic KL scores are underway. The outcome of this work will provide an indication of the performance of this approach in assessing OA severity compared with the current 'gold standard' radiographic assessment. Following further validation of these novel technologies through large-scale prospective studies, they may be developed to support stratified approaches for designing clinical trials of new treatments and interventions as well as providing earlier outcome measures for use in such studies.

In clinical practice they may offer the potential to develop approaches for early identification of people at risk of progressing-to-severe knee OA, and to develop and deliver personalized treatment approaches, as well as to monitor the efficacy of interventions. Furthermore, both approaches have potential for use in other affected joints. The extent to which 3D MRI and AE will qualify for use as biomarkers in these, and other, applications remains to be determined. Nevertheless, the fact that this, and other work in this field, is progressing promisingly suggests that robust biomarkers for knee OA are at last becoming a feasible prospect.

Financial & competing interests disclosure

J Goodacre, L Shark and P Diggle are supported by an MRC DPFS grant, reference MR/K025597/1. M Bowes is a shareholder and employee of Imorphics Ltd. J Waterton holds stock options in AstraZeneca, a for-profit company engaged in the discovery, development, manufacture and marketing of proprietary therapeutics. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed. No writing assistance was utilized in the production of this manuscript.

Executive summary

- While the paucity of biomarkers for knee osteoarthritis continues to constrain progress in developing new treatments for this condition, several promising approaches for biomarker development are beginning to emerge.
- New approaches and technologies being used for biomarker development in knee osteoarthritis also have potential to inform new paradigms about this condition for application in clinical practice and research.

References

Papers of special note have been highlighted as:
•• of considerable interest.

1 National Institute for Heath and Care Excellence. Osteoarthritis: care and management in adults. (2014). www.nice.org.uk/guidance/cg177/resources/guidance-osteoarthritis-pdf

•• **Authoritative description of current principles for clinical management of osteoarthritis in the UK.**

2 Hunter DJ, Guermazi A, Roemer F, Zhang Y, Neogi T. Structural correlates of pain in joints with osteoarthritis. *Osteoarthritis Cartilage* 21(9), 1170–1178 (2013).

3 Hill CL, Gale DG, Chaisson CE *et al.* Knee effusions, popliteal cysts, and synovial thickening: association with knee pain in osteoarthritis. *J. Rheumatol.* 28(6), 1330–1337 (2001).

4 Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin. Pharmacol. Ther.* 69(3), 89–95 (2001).

5 Lesko LJ, Atkinson AJ Jr. Use of biomarkers and surrogate endpoints in drug development and regulatory decision making: criteria, validation, strategies. *Annu Rev. Pharmacol. Toxicol.* 41, 347–366 (2001).

6 Hunter DJ, Guermazi A, Lo GH *et al.* Evolution of semi-quantitative whole joint assessment of knee OA: MOAKS (MRI Osteoarthritis Knee Score). *Osteoarthritis Cartilage* 19(8), 990–1002 (2011).

7 Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. *Stat. Med.* 8(4), 431–440 (1989).

8 Wagner JA. Strategic approach to fit-for-purpose biomarkers in drug development. *Annu. Rev. Pharmacol. Toxicol.* 48, 631–651 (2008).

9 Waterton JC. Chapter 12 translational magnetic resonance imaging and spectroscopy: opportunities and challenges. In: *New Applications Of NMR In Drug Discovery And Development*. Garrido L, Beckmann N (Eds). The Royal Society of Chemistry, 333–360 (2013).

10 Hill AB. The environment and disease: association or causation? *Proc. R. Soc. Med.* 58, 295–300 (1965).

11 Chetty RK, Ozer JS, Lanevschi A *et al.* A systematic approach to preclinical and clinical safety biomarker qualification incorporating Bradford Hill's principles of causality association. *Clin. Pharmacol. Ther.* 88(2), 260–262 (2010).

12 Hoefield RA, Kalra PA, Baker P *et al.* Factors associated with kidney disease progression and mortality in a referred CKD population. *Am. J. Kidney Dis.* 56(6), 1072–1081 (2010).

13 Levey AS, Coresh J. Chronic kidney disease. *Lancet* 379(9811), 165–180 (2012).

14 Buyse M, Molenberghs G. Criteria for the validation of surrogate endpoints in randomized experiments. *Biometrics* 54(3), 1014–1029 (1998).

15 Psaty BM, Weiss NS, Furberg CD *et al.* Surrogate end points, health outcomes, and the drug-approval process for the treatment of risk factors for cardiovascular disease. *JAMA* 282(8), 786–790 (1999).

16 Mascaro B, Prior J, Shark LK, Selfe J, Cole P, Goodacre J. Exploratory study of a non-invasive method based on acoustic emission for assessing the dynamic integrity of knee joints. *Med. Eng. Phys.* 31(8), 1013–1022 (2009).

•• **Describes early results from developing the acoustic emission technique and protocol.**

17 Prior J, Mascaro B, Shark LK *et al.* Analysis of high frequency acoustic emission signals as a new approach for assessing knee osteoarthritis. *Ann. Rheum. Dis.* 69(5), 929–930 (2010).

18 Shark LK, Chen H, Goodacre J. Knee acoustic emission: a potential biomarker for quantitative assessment of joint ageing and degeneration. *Med. Eng. Phys.* 33(5), 534–545 (2011).

•• **Describes further development of acoustic emission measurement and data analysis methods.**

19 Shark LK, Chen H, Goodacre J. Discovering differences in acoustic emission between healthy and osteoarthritic knees using a four-phase model of sit-stand-sit movements. *Open Med. Inform. J* 4, 116–125 (2010).

20 Eckstein F, Cicuttini F, Raynauld JP, Waterton JC, Peterfy C. Magnetic resonance imaging (MRI) of articular cartilage in knee osteoarthritis (OA): morphological assessment. *Osteoarthritis Cartilage* 14(Suppl. A), A46–A75 (2006).

21 Peterfy C, Woodworth T, Altman R. Workshop for consensus on osteoarthritis imaging: MRI of the knee. *Osteoarthritis Cartilage* 14, 44–45 (2006).

22 Sowers MF, Hayes C, Jamadar D *et al.* Magnetic resonance-detected subchondral bone marrow and cartilage defect characteristics associated with pain and X-ray-defined knee osteoarthritis. *Osteoarthritis Cartilage* 11(6), 387–393 (2003).

23 Link TM, Steinbach LS, Ghosh S *et al.* Osteoarthritis: MR imaging findings in different stages of disease and correlation with clinical findings. *Radiology* 226(2), 373–381 (2003).

24 Wluka AE, Ding C, Jones G, Cicuttini FM. The clinical correlates of articular cartilage defects in symptomatic knee osteoarthritis: a prospective study. *Rheumatology (Oxford)* 44(10), 1311–1316 (2005).

25 Wildi LM, Raynauld JP, Martel-Pelletier J, Abram F, Dorais M, Pelletier JP. Relationship between bone marrow lesions, cartilage loss and pain in knee osteoarthritis: results from a randomised controlled clinical trial using MRI. *Ann. Rheum. Dis.* 69(12), 2118–2124 (2010).

26 Eckstein F, Guermazi A, Gold G *et al.* Imaging of cartilage and bone: promises and pitfalls in clinical trials of osteoarthritis. *Osteoarthritis Cartilage* 22(10), 1516–1532 (2014).

27 Hunter DJ, Zhang W, Conaghan PG *et al.* Responsiveness and reliability of MRI in knee osteoarthritis: a meta-analysis of published evidence. *Osteoarthritis Cartilage* 19(5), 589–605 (2011).

28 Bowes MA, Vincent GR, Wolstenholme CB, Conaghan PG. A novel method for bone area measurement provides new insights into osteoarthritis and its progression. *Ann. Rheum. Dis.* 74(3), 519–525 (2015).

•• **Describes advanced method for bone area measurement in knee magnetic resonance images.**

29 Wang Y, Wluka AE, Cicuttini FM. The determinants of change in tibial plateau bone area in osteoarthritic knees: a cohort study. *Arthritis Res. Ther.* 7(3), R687–R693 (2005).

30 Eckstein F, Hudelmaier M, Cahue S, Marshall M, Sharma L. Medial-to-lateral ratio of tibiofemoral subchondral bone area is adapted to alignment and mechanical load. *Calcif. Tissue Int.* 84(3), 186–194 (2009).

31 Wells JG, Rawlings RD. Acoustic emission and mechanical properties of trabecular bone. *Biomaterials* 6(4), 218–224 (1985).

32 Singh VR. Acoustical imaging techniques for bone studies. *Applied Acoustics* 27(2), 119–128 (1989).

33 Schwalbe HJ, Bamfaste G, Franke RP. Non-destructive and non-invasive observation of friction and wear of human joints and of fracture initiation by acoustic emission. *Proc. Inst. Mech. Eng. H* 213(1), 41–48 (1999).

34 Rajachar RM, Chow DL, Curtis CE, Weissman NA, Kohn DH. Use of acoustic emission to characterize focal and diffuse microdamage in bone. In: *Acoustic Emission:*

Standards and Technology Update. STP 1353. Vahaviolos, SJ (Ed.). (1999).

35 Watanabe Y, Takai S, Arai Y, Yoshino N, Hirasawa Y. Prediction of mechanical properties of healing fractures using acoustic emission. *J. Orthop. Res.* 19(4), 548–553 (2001).

36 Cardoso L, Teboul F, Sedel L, Oddou C, Meunier A. *In vitro* acoustic waves propagation in human and bovine cancellous bone. *J. Bone Miner. Res.* 18(10), 1803–1812 (2003).

37 Tatarinov A, Sarvazyan N, Sarvazyan A. Use of multiple acoustic wave modes for assessment of long bones: model study. *Ultrasonics* 43(8), 672–680 (2005).

38 Shrivastava S, Prakash R. Assessment of bone condition by acoustic emission technique: a review. *J. Biomed. Sci. Eng.* 2(3), 144–154 (2009).

39 Lu Q, Yadid-Pecht O, Sadowski D, Mintchev MP. Acoustic and intraluminal ultrasonic technologies in the diagnosis of diseases in gastrointestinal tract: a review. *Engineering* 5(5B), 73–77 (2013).

40 Cho NY, Ferracane JL, Lee IB. Acoustic emission analysis of tooth-composite interfacial debonding. *J. Dent. Res.* 92(1), 76–81 (2013).

41 Choi N-S, Gu J-U, Arakawa K. Acoustic emission characterization of the marginal disintegration of dental composite restoration. *Composites Part A: Applied Science and Manufacturing* 42(6), 604–611 (2011).