Validating markers in osteoarthritis

David T Felson

Boston University School of Medicine, The Arthritis Center, 80 East Concord Street, A203, Boston, MA 02118, USA. Tel + 617-638-5180. Fax + 617-638-5239.

This conference focuses on exciting new work evaluating molecular markers for joint and skeletal diseases. Yet the term "markers" has not been operationally defined. Further, while many studies are now attempting to evaluate the correlation between synovial fluid or serum "markers" and the prevalence or development of such diseases, no one has yet laid out "Koch's postulates" for markers, the criteria that would be necessary to prove the existence of a marker for disease. I shall try both to discuss the different meanings of the word marker when used in the context of a biochemical measure of osteoarthritis and try to develop criteria that might be used to validate whether a potential marker should be accepted definitively.

"Markers" and their potential meaning

First, I shall not try to define the word "marker", since the word has been used in many different contexts. Cytokines, enzymes and their inhibitors, matrix components and their fragments, antibodies to cartilage collagen and membrane proteins of chondrocytes and even growth factors have all been characterized as markers for osteoarthritis. These molecules could serve as measures of disease severity, measures of cartilage or bone turnover or risk factors for disease development. In reality there are different types of markers (Table 1), just as in other branches of medicine, there are different types of diagnostic and prognostic tests.

On the one hand, markers can serve as diagnostic tests helping to distinguish OA from another joint disease. Examples of serologically measured diagnostic tests include rheumatoid factor and ANA. The synovial fluid leukocyte count can also serve as a diagnostic test.

The term "marker" also might be used for a test evaluating disease severity rather than its presence or absence. In chronic obstructive lung disease, the forced vital capacity and FEV1 (the amount of breath expired in the first second of expiration) are both measures of disease severity. TNM classifications for cancers are measures of disease severity rather than diagnostic markers. Some biochemical tests like TSH serve both as diagnostic tests (e.g. for hypothyroidism) and as measures of disease severity. In OA, disease severity is measured by Kellgren and Lawrence grade, by the amount of cartilage loss on arthroscopy or by the degree of functional impairment experienced.

Often, biochemical assays developed to evaluate osteoarthritis have been touted as prognostic markers and tested to see whether they predict the later occurrence or worsening of osteoarthritis. Prognostic markers fit into two different categories, disease severity markers (mentioned above) and measures of disease dynamics. Disease severity markers simply evaluate the extent of disease (see above). Measures of disease dynamics evaluate the ongoing repair and degradative processes occurring within a joint. In osteoporosis, measures of disease dynamics include osteocalcin, a marker of osteoblast function.

Other types of markers are also used in medicine, and similar osteoarthritis markers could also be envisioned. One type of marker is a predictor of response to therapy, such as estrogen receptors for breast cancer. One could, for example, envision a marker of collagenase activity that would identify osteoarthritis

Table 1. Different types of markers

1. Diagnostic tests (e.g. rheumatoid factor, ANA)
2. Measures of disease severity (e.g. FEV1 in COPD, TNM in cancer)
3. Prognostic markers
a. Disease severity markers e.g TNM, FEV1, Prothrombin time in cirrhosis
b. Measures of disease dynamics (osteocalcin in osteoporosis)
4. Predictors of responsiveness to treatment (e.g. estrogen receptors in breast cancer)
5. Measures of response to treatment (e.g. thyroid function tests, CEA)

subjects who are most likely to respond to a collagenase inhibitor. Another type of marker might be a measure of response to treatment. In medicine examples include thyroid function tests and for cancer, the carcinoembryonic antigen (CEA), which rises on disease recurrence in many cancers.

While a number of biochemical moieties have been identified that are correlated with osteoarthritis and may even predict the likelihood of disease progression, some of these are not really markers. They are risk factors for disease. Risk factors affect disease pathogenesis or affect the repair process within the joint. They are not measures of the actual repair or degradative process, nor are they measures of actual disease severity, such as cartilage loss. Although we do not necessarily understand whether growth factors reflect repair and degradative processes, it is likely that factors such as IGF1 are appropriately viewed as protective risk factors for disease rather than markers of disease severity or prognosis. Cytokines that influence the levels of degradative enzymes within cartilage might also be viewed as risk factors either potentiating or developing the ongoing osteoarthritic process. Genetic predisposition to osteoarthritis probably should not be viewed as a marker but rather as a risk factor for accelerated or early onset disease. That is not to diminish the critical importance of all of these factors in our understanding of osteoarthritis pathogenesis. It suggests that we have been using the word "marker" too liberally.

Validity and reliability

To determine whether a marker is an accurate measure of an underlying disease, we must evaluate its validity. Validity is defined as "the extent to which any instrument measures what it is intended to measure" (Carmines and Zeller). Although there are many different types of validity that can be tested, one central type of validity of interest for markers is criterion validity. For biological phenomena, criterion validity can be considered as the test of whether a marker measures the ongoing biological event it is purported to measure. One must assume that there is a gold standard measure of the underlying event, such as an imaging procedure which permits one to evaluate whether a person has osteoarthritis or permits an assessment of the extent of osteoarthritis.

Obviously, measures against which markers ought to correlate include radiographic severity of disease, severity of disease as documented by other imaging techniques like MRI or arthroscopy, the likelihood of the patient needing a total joint replacement or other joint surgery, functional impairment and pain in the joint.

Table 2. Modified Koch's postulates' for biochemical markers of the state of OA

1. Marker is biologically credible
2. Marker is regularly (universally) found in patients with osteoarthritis of a joint
3. Normal values for marker by age defined sensibly
4. With change in state of OA, marker changes appropriately.

There are two other types of validity relevant to the development of markers for OA. Content validity reflects the extent to which an instrument (or marker) measures the range of disease expression. An OA marker might have adequate content validity if it correlated with the severity of disease across the range of OA, from early disease to late, regardless of the joint affected.

Face validity is the extent to which the marker is believable. This reflects the credibility of the underlying scientific work performed prior to testing the marker as a clinical tool.

Validating markers

Markers as diagnostic tests

If a biochemical marker for osteoarthritis is being validated as a diagnostic test for disease, it should fulfill certain criteria (Table 2). Koch developed a set of postulates that served as standards for proof that microorganisms caused a particular disease; a modified set of Koch's postulates could be used to prove that a biochemical marker is a valid test for osteoarthritis. First, the marker should be biologically credible (it should have face validity). Second, Koch required that a microorganism be found regularly (universally) in an infected person in order that the microorganism be tied causally to disease. It is probably insufficient for a biochemical marker of osteoarthritis to be found significantly more often in osteoarthritis than in normal subjects. To be proven as a marker for diagnosis, such a marker ought to be found universally (or nearly universally) in all of those with osteoarthritis. Serum levels of a marker might be spuriously normal if only one joint were afflicted by OA and an insufficient quantity of the marker were released from the joint. Nonetheless, if joint synovial fluid or even cartilage levels of a marker are available, then one should be able to detect abnormal levels of marker in nearly everyone with disease. In the same sense, normal values should be

Table 3. Suggested criteria for validating a marker of OA prognosis—all are necessary

1. Strong biological rationale
   - Marker identifiable in cartilage or periarticular bone
   - If measure of O.A. dynamics, its role in pathogenesis of OA at least partly understood
   - If measure of disease severity, must correlate with disease severity as determined by imaging and/or functional assessment
2. Marker measured at baseline in a body fluid predicts the course of OA. Course of OA at baseline and follow-up is determined independently of marker measurement using conventional clinical means
3. Marker validated in patients with spectrum of mild\severe disease and OA of different etiologies (eg. post-traumatic vs primary)
4. Marker measurement is reliable (repeatable) and described in sufficient detail so as to be replicable by others.

defined by age, and persons without disease should have measurable levels of markers which are within the normal range. Lastly, with a change in the state of osteoarthritis defined independently of the marker, there should be an appropriate change in the level of the marker.

Markers for OA prognosis

Many, if not most, of the putative markers for osteoarthritis are not necessarily markers of the existence or nonexistence of osteoarthritis but are rather markers of disease dynamics or disease severity. As noted earlier, these are often characterized as prognostic markers. To validate prognostic markers, suggested criteria are shown in Table 3; each is necessary but not sufficient for validation.

First, a prognostic marker should have a strong biological rationale (face validity). A marker should be identifiable in cartilage or periarticular bone. If it is a measure of disease severity, it must correlate with disease severity as determined independently by an imaging measure and/or measure of joint dysfunction. A prognostic marker should correlate strongly with the course of osteoarthritis in an individual. If the level of a marker is abnormal at baseline, then the risk of subsequent joint deterioration should be magnified (predictive validity, a type of criterion validity). The evaluation of the course of osteoarthritis should be made by means that are conventionally accepted and are independent of the marker.

To be an adequate prognostic marker, the marker should be valid in those with different levels of disease and disease of different etiologies (content validity).

Lastly, separate from validity, any biochemical measure should be repeatable and described in sufficient detail so as to be replicable by others.

Validating other types of markers

As noted earlier, markers could serve yet other functions such as predictors of responsiveness to treatment. A putative marker should be assayed at baseline in a longitudinal treatment study and those who respond to treatment, defined independently of the marker, should have a different marker level at baseline than those who did not respond to treatment.

Measure of response to treatment is slightly different. For a marker of treatment response, measures at baseline and after treatment should be obtained. Normalization or improvement in the level of marker should occur in those who respond to treatment.

Conclusion

Biochemical markers for osteoarthritis are multifaceted and could serve different diagnostic purposes. Some serve as measures of disease severity, others as measures of prognosis and yet others, as measures of responsiveness to treatment. Obviously, some markers may serve multiple functions. To function as adequate tests, they should meet a set of standards which depend on whether they are being tested as diagnostic tests, prognostic tests, etc. It is only when these markers have met such criteria that they will be generally accepted in the clinical community and will become widely used.

Acknowledgements

Supported by NIH Grants AR20613 and AG09300

Reference

Carmines E G, Zeller R A. Reliability and Validity Assessment. Beverly Hills, Sage Publications, Inc., 1979: pp. 9–70.