



Published in final edited form as:

Curr Drug Targets. 2010 May ; 11(5): 536–545.

A Pathway and Approach to Biomarker Validation and Qualification for Osteoarthritis Clinical Trials

David J. Hunter^{1,*}, Elena Losina², Ali Guermazi³, Deb Burstein⁴, Marissa N. Lassere⁵, and Virginia Kraus⁶

¹Division of Research, New England Baptist Hospital, 125 Parker Hill Ave, Boston MA 02120, USA

²Orthopedics and Arthritis Center for Outcomes Research Brigham and Women's Hospital, 75 Francis Street, Boston, MA, USA

³Boston University Medical Center, 820 Harrison Avenue, FGH Building, 3rd Floor, Boston, MA 02118, USA

⁴Department of Radiology, Beth Israel Deaconess Medical Center, 4 Blackfan Circle, Boston, MA 02115, USA

⁵Department of Rheumatology, St George Hospital, Gray Street, Kogarah, Sydney 2217, NSW, Australia

⁶Duke University, GSRB1 Bldg, Room 1033A, 595 La Salle St., Durham, NC 27710, USA

Abstract

This narrative review outlines the work done in other fields with regards biomarker validation and qualification and the lessons that we may learn from this experience. Defining a universally agreed upon path for biomarker validation and qualification is urgently needed to circumvent many of the hurdles faced in OA therapeutic development irrespective of whether we are discussing biochemical markers, imaging markers or other measures. This review proposes a path that may be suitable for osteoarthritis and poses some logical next steps that will take us in this direction.

Keywords

Osteoarthritis; biomarkers; imaging; biochemical marker; validation; qualification; intervention; surrogate

INTRODUCTION

Studies with hard clinical endpoints are not always the first ones possible for establishing the efficacy or risk of an intervention [1]. It is common practice in the development of interventions to use proxies for clinical outcomes using biomarkers as intermediary mechanisms of disease processes. Such biomarkers are defined as characteristics that are measured as an indicator of a biological or pathological process, or of a pharmacological

© 2010 Bentham Science Publishers Ltd.

*Address correspondence to this author at the Division of Research, New England Baptist Hospital, 125 Parker Hill Ave, Boston MA 02120, USA; Tel: 617 754 6655, Fax: 617 754 5728; djhunter@caregroup.harvard.edu.

DISCLOSURE

This is a narrative review and the comments and editorial expressed herein represent those of the author/s and do not reflect those of any official scientific role or institution that the author/s may be hold or be affiliated with.

response [2]. In clinical trials biomarkers in addition to allowing close monitoring of response to treatment, may enable selection of patients most likely to respond to specific therapies. Some biomarkers may be useful in identifying early signs of toxicity [3].

The high level of interest in biomarkers research is governed by increased research and development costs, and the ultimate need for mechanisms describing scientific advance and therapeutic breakthrough. The time and cost needed to develop a new compound have further increased in recent years [4]. DiMasi *et al.* calculated that the average cost of bringing a drug to the market increased from US \$ 318 million in 1991 to US \$ 802 million in 2003 (inflation adjusted, including opportunity cost of capital) [5]. The cost calculation comprises the expenses for failures of drug candidates in the development process. The average probability that a drug candidate will successfully pass clinical phase I studies is in the range of 75%; the respective values for phases II and III trials are 50% and 65% [6]. In total (including further probabilities, e.g. for the regulatory review), the cumulative probability that a leading drug candidate will successfully proceed from the preclinical phase to approval is about 8% (i.e. for every 12–13 compounds that were serious candidates in preclinical research, only one drug will make it onto the market) [6]. The rising cost of drug development is imposing a significant burden on industry engaged in therapeutic development. The attraction of integrating biomarkers into the therapeutic development process includes the expectation that less promising projects may be stopped earlier (especially before they enter into costly clinical phase III [7]) and that the total cost of drug development will be optimized.

Historically disease knowledge development and treatment innovation in osteoarthritis (OA) has been considered to be slow. One of the many reasons purported as responsible for this slow pace has been the alleged lack of valid and responsive biomarkers to ascertain efficacy, which itself has been dependent upon the slow evolution of the understanding of the complex nature of joint tissue biology. The apparent lack of a transparent pathway for biomarker validation and qualification has been perceived as a barrier to a faster pace of discovery. Defining a universally agreed upon path for biomarker validation and qualification is urgently needed to circumvent many of the hurdles faced in OA therapeutic development irrespective of whether we are discussing biochemical markers, imaging markers or other measures [4].

The multiple and complex hurdles faced in OA may ultimately be responsible for the slow pace of therapeutic development. Some hurdles that are somewhat unique to OA biomarker development that compound this situation include the following factors. First, our current reference standard for disease diagnosis and severity is often the radiograph, which has a low responsiveness to change and at most moderately correlates with clinical endpoints. Second, there is a lack of consensus for surrogate measure and efficacy of intervention development and definition as to what constitutes a meaningful clinical endpoint. Third, OA is extraordinarily complex with marked heterogeneity in onset, clinical presentation, rate of disease progression, pattern of joint involvement and synovial tissue structure affected. Thirdly, there is no clear consensus of a pathway for biomarker validation and this uncertainty has stalled many development programs. This vicious cycle of imperfect biomarkers to test efficacy of disease modifying therapies in clinical trials and the lack of effective therapies to demonstrate the validity of biomarkers has challenged therapeutic development for years. What remains clear however is that it creates exciting opportunities to refine existing biomarker methods and identify new ones for accelerating the development of safe and effective treatments for OA.

This narrative review outlines the work done in other fields with regards biomarker validation and qualification and the lessons that we may learn from this experience. It then

proposes a pathway/approach to biomarker validation and qualification for OA clinical trials. It is the intent of this review to propose a method that could be used for all biomarkers used in OA. The most common biomarkers currently tested to be used in OA are biochemical and imaging markers however the pathway proposed should not be seen as limited to these biomarkers. The focus of this review is on OA clinical trials and the validation strategies proposed will focus on the efficacy of intervention markers and more specifically the role of surrogate markers in defining the efficacy of intervention. This review will summarize current biomarkers definitions and classification of biomarkers, provide overview of the validation techniques and lay out overarching principles of biomarkers qualification and conclude with outlining challenges in biomarker discovery in OA and suggest a means of overcoming these challenges.

COMMON DEFINITIONS

It is critical to delineate what we mean by the various terms used, as current usage is often incorrect, and this ambiguity may stem from an incorrect understanding of appropriate definitions [8, 9]. Whilst there are several definitions that have been proposed [2,10–13] the brief synthesis of some working definitions is as follows:

1. **biological marker** (biomarker)— a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes or pharmacologic responses to a therapeutic agent.
2. **surrogate marker or endpoint**—a biomarker that is intended to serve as a substitute for a clinically meaningful endpoint and is expected to predict the effect of a therapeutic intervention; and
3. **clinical endpoint**—a clinically meaningful measure of how a patient feels, functions, or survives.

The hierarchical distinction between biomarkers and surrogate endpoints is intended to indicate that relatively few biomarkers will meet the stringent criteria that are needed for them to serve as reliable substitutes for clinical endpoints [14]. Many authors have published their validation criteria for surrogate endpoints. The most rigorous standards are those of Fleming & DeMets [15], who stipulated that both of the following conditions must be satisfied: (a) The surrogate endpoint must be correlated with the true clinical outcome; and (b) as initially proposed by Prentice [16], the surrogate endpoint must fully capture the net effect of treatment on clinical outcome.

A **valid biomarker** is defined as “a biomarker that is measured in an analytical test system with well-established performance characteristics and for which there is an established scientific framework or body of evidence that elucidates the physiologic, toxicologic, pharmacologic, or clinical significance of the test results [17]”. The validity of a biomarker is closely linked to what we think we can do with it. This biomarker context drives not only how we define a biomarker but also the complexity of its qualification.

Validation is the process of assessing the biomarker and its measurement performance characteristics, and determining the range of conditions under which the biomarker will give reproducible and accurate data [2, 18]. The evidentiary process of proving a linkage between the biomarker and a clinical end point was termed ‘evaluation’ in preference to validation. More recently, evaluation has been replaced with qualification, which has become accepted terminology [2]. **Qualification** is the evidentiary process of linking a biomarker with biological processes and clinical end points [2, 19]. The biomarker literature occasionally uses “validation” and “qualification” or “evaluation” interchangeably. We have avoided this

because the validation and qualification processes must be distinguished, and the term “validation” does not adequately describe the qualification process.

Next section provides more detail overview of validation process.

BIOMARKER VALIDATION

Validation of a biomarker is a necessary component to delivery of high-quality research data necessary for effective use of biomarkers. Biomarkers pass through three evidentiary stages towards full acceptance under regulatory guidance: exploratory, probable valid and known valid [17, 20]. More recently in the validation framework the terminology for “known valid” has been replaced by a new term of “fit-for-purpose”. The importance of all these above definitions, from a method validation perspective, is that the further along the spectrum towards a surrogate end point the biomarker is positioned the greater the degree of thoroughness necessary to validate the biomarker assay. Under this schema qualification [10] could ensue once adequate validation was complete.

The criteria for validation are defined by the nature of the question that the biomarker is intended to address, the degree of certainty that is required for the answer, and the assumptions about the relationship between changes in the biomarker and clinical endpoints [21]. Validation has been described as not being an all-or-none (binomial) variable, such as the outcome of an efficacy trial, but a continuous variable that varies during the drug development process as new information and data are obtained. As described below there are a number of distinct criteria for validation which can be completely or partially met; hence validation is a multi-step process and within those steps the strength of validation can vary.

The scientific program for evaluating biomarkers must be planned as early as possible in the therapeutic discovery and preclinical period of therapeutic development with a blueprint to bring that biomarker into clinical trials and to establish the link between the biomarker and the clinical outcome. There are multiple dimensions to biomarker validation that encompass important elements of study design and data analysis, including statistical assessment. There are also multiple pathways to validation of a biomarker for an intended use, and validation data itself is likely to arise from the totality of evidence provided progressively by preclinical animal studies, early Phase I and Phase II clinical studies in healthy volunteers or patients, and late-phase efficacy and safety trials in patients with the targeted disease.

Typically, validation takes into account the following properties of a biomarker using the following criteria [12]:

1. *Sensitivity of the biomarker*, referred to as the ability of an appropriate biomarker or a change in biomarker to be measured with adequate precision, and with sufficient magnitude of change, to make it sensitive enough to reflect a meaningful change in important clinical endpoints. Sensitivity also describes the quality of the relationship between the magnitude of change in the biomarker and the magnitude of change in the clinical endpoint because a high level of correlation, does not necessarily prove a cause-effect relationship. For example, in a clinical trial of an anti-resorptive for osteoporosis, bone mineral density and its change may be sensitive indicator related to the clinical endpoint of fracture, however there are other factors such as falls risk that can also influence fracture risk.
2. *Specificity of the biomarker*, referred to as the ability of a biomarker or a change in biomarker to distinguish patients who are responders to an intervention from those who are non-responders in terms of changes in clinical endpoints. Specificity defines the extent to which a biomarker explains all or most of the changes in a

clinical endpoint and can be used for both categorical and continuous endpoints. As an example that is pertinent to OA, the recent trial of risedronate demonstrated effects on a marker of type II collagen metabolism in the drug group that would be consistent with slowing progression, however this effect was not reflected in a similar slowing of the rate of progression as measured by x-ray. This suggests that this biomarker may not have good specificity on reflecting JSN progression under the influence of risedronate [7].

3. *Bioanalytical assessment of the laboratory* or test measurement of the biomarker in terms of accuracy, precision, reproducibility, range of use, limit of detection, and variability.
4. *Probability of false positives*, defined by situations in which a desired change in a biomarker is not reflected by a positive change in a clinical endpoint or, even worse, is associated with a negative change in a clinical endpoint. A hypothetical example would be the use of a type 2 collagen breakdown marker to predict incident OA development. In this example a false positive was the detection of elevated levels of the biochemical marker in the absence of incident OA development.
5. *Probability of false negatives*, defined by situations in which no change or a small observed change in a biomarker fails to signal a positive, meaningful change in a clinical endpoint. A hypothetical example consistent with the use described in point 4 above would be the use of a type 2 collagen breakdown marker that failed to predict persons at risk of incident OA.
6. A PK-PD (Pharmacokinetic/Pharmacodynamic) model that has been shown to predict future clinical outcomes or suitable dose adjustments based on biomarker measurement. This establishes the correlation between changes in the biomarker and changes in drug exposure, measured as plasma concentration or dose. One of the challenges here is to prospectively plan and properly implement the model and to determine which metrics of drug exposure and biomarker time course are best able to predict clinical outcomes.

Alternative set of validation criteria has been proposed by OMERACT (Outcome measures for rheumatoid arthritis clinical trials), an informal collaborative group of professionals dedicated to improving outcome measurement in rheumatic disease and is called the OMERACT filter [22]. The OMERACT process is data-driven, iterative and consensus building. This extends, and provides some overlap with the concepts of validation just mentioned and affords some precedent of acceptance in the rheumatologic community. The criterion of the OMERACT filter include [22, 23]:

1. **Truth:** is the measure truthful, does it measure what it intends to measure? Is the result unbiased and relevant? This criterion captures the issues of face, content, construct and criterion validity. Accuracy and precision, terms commonly used in the biomedical literature, also capture truth, but they are not synonymous. For example, in a method-comparison study of continuous measures, to determine how well the measures agree we calculate the difference between the two measurements for each subject. Accuracy is the mean of these differences whereas precision is the standard deviation of the differences. However, precision has also been used to describe the property of reliability or consistency in medicine and this confusion in terminology has caused problems both with the design of method-comparison studies and their analysis.
2. **Discrimination:** captures the issues of reliability and sensitivity to change (also known as responsiveness or discriminant validity). Does the measure discriminate

between situations that are of interest? The situations can be states at one time (for classification or prognosis) or states at different times (to measure change).

- 3. Feasibility:** can the measure be applied easily, given constraints of time, money, and interpretability? This criterion addresses the pragmatic reality of the use of the measure, one that may be decisive in determining a measure's success.

Under this rubric it is worth considering that the process of validation permits the simultaneous examination of multiple aspects of measurement performance. Whilst the clinimetric properties contained in truth (validity) and discrimination (responsiveness and reliability) are consistent with many other recommendations with regards biomarker validation, one additional advantageous element afforded in the OMERACT filter is consideration of the feasibility. Critically aspects such as brevity, simplicity, costs, availability, participant burden are critical concepts for considering the application of biomarkers in both research and clinical practice. The template of the OMERACT filter provides a useful data driven iterative approach for evaluating an outcome measure [22].

The most desirable paradigm for validation of biomarkers is provided by adequate and well-controlled clinical studies that (a) define standardized relationships between therapeutic exposure and response, (b) test hypotheses regarding mechanism of therapeutic action, and (c) provide estimates of the magnitude of benefit. The size and duration of the treatment effect are essential aspects of biomarker evaluation, but sample size and study design are also important [12].

The process of *surrogate endpoint validation* (a sub-type of efficacy biomarker) is a much larger hurdle [24, 25]. To validate a biomarker as a surrogate endpoint Ross Prentice identified two conditions that, if simultaneously valid, would be sufficient: (1) The biological marker must be correlated with the clinical endpoint; and (2) the marker must fully capture the net effect of the intervention on the clinical-efficacy endpoint [16]. Although many have had the misunderstanding that the first condition would be adequate to validate a surrogate, the second required condition is less likely to be satisfied and is much more difficult to verify [14]. For example, in the treatment of osteoporosis using anti-resorptive agents there is interest in quantifying the relationship between fracture endpoints and surrogates such as bone mineral density (BMD) or bone turnover markers [26]. However analyses based on individual patient data report that only a limited proportion of the anti-fracture efficacy is explained by BMD increases for agents such as alendronate, risedronate and raloxifene have given reason for concern over whether BMD is actually a true surrogate. Whilst the actual BMD value is correlated with fracture risk and thus BMD is useful in identifying patients that might need treatment, there is limited evidence to support BMD increase with anti-resorptive agents as a reliable substitute for fracture risk reduction. Other approaches to surrogate endpoint validation are summarized elsewhere [12, 27–30].

Validation of a surrogate marker or endpoint should be based on both in-depth clinical insights and empirical evidence. Ideally, one should have a comprehensive understanding of the causal pathways of the disease process and of the intervention's unintended and intended mechanisms of action. Admittedly, achieving such understanding is an extremely complicated challenge. Hence, as recognized by several researchers, validation of a potential surrogate endpoint typically also requires a meta-analysis of many RCTs [15]. As a result, it is easier to directly show the effect of an intervention on the clinical-efficacy endpoint than to actually validate the surrogate. Once a surrogate is "validated" for one pharmacologic class of treatment regimens (such as MMP inhibitors), it is tempting to consider that it can be validly used as a replacement endpoint when evaluating other classes of agents (such as bisphosphonates) as well. However, one must be able to conclude that the "alternative beneficial effects" and "unintended negative effects" on the clinical-efficacy outcome that

are not directly captured by the surrogate endpoint will yield the same net effect for the other classes of agents as for the class of agents used in the validation analyses. Thus the validation of a surrogate endpoint is typically for a given disease setting (e.g. knee vs. hip vs. hand OA) and for the class of agents studied in those clinical trials. Surrogate endpoints which have been accepted by some regulatory authorities for drug approval purposes in the (accelerated) approval context include the RECIST (Response Evaluation Criteria In Solid Tumors) criteria for tumor response assessment in oncological trials (surrogate for the clinical endpoint: survival), lowering of cholesterol levels (clinical endpoint: cardiovascular events), increase in CD4 count (clinical endpoint: improved survival in AIDS) and the number of cerebral lesions on MR imaging (clinical endpoint: disease progression in multiple sclerosis) [4].

While various statistical approaches have been proposed, it appears that all surrogate validation methods focus on the following three requirements [26, 31].

1. A valid surrogate must be correlated with the clinical endpoint.
2. A valid surrogate should capture a reliable and sufficiently large portion of the treatment effect on the clinical endpoint.
3. A valid surrogate should be able to predict the treatment effect on the clinical endpoint.

Satisfaction of these three requirements requires distinct statistical approaches [31].

Alternate methods have been also been proposed for the validation of genomic biomarkers [32].

BIOMARKER QUALIFICATION

Historically biomarker validation was settled by debate, consensus, and the passage of time. Although intellectually painless this process was slow [20]. The need to accelerate this process has prompted regulatory guidance on mechanisms to do so. The historical and traditional absence of a structured qualification process for biomarkers promoted a lack of transparency, understanding and further promoted complexity [10]. In contrast, the recent development of a strategic pathway for biomarker qualification provides a uniform, consistent method for advancing specific biomarkers for specific contexts. The explicit methods of validation vary depending on the process mapped out however recent regulatory guidance on qualification affords robustness and clarity to what was once an unclear process.

Like validation, a number of different strategies for biomarker qualification have been proposed [10, 19], some of which are context dependent and others context independent (for example the methods to replace animal testing in toxicology and markers of toxicity). Context and qualification for new biomarkers is assessed relative to current biomarkers. If the measurement performance of current biomarkers is not perfect relative to a specific end point, the context and qualification of new biomarkers may not be accurately established.

A high level of stringency is required when a biomarker response is substituted for a clinical outcome and is proposed as the basis for regulatory approval of an application to market a new drug. Thus understandably the Food and Drug Administration (FDA) has a number of position papers delineating acceptable standards for biomarker validation and acceptance. Steps towards biomarker qualification are clearly delineated by regulators lead by the FDA [10].

The FDA biomarker qualification process was designed around the Interdisciplinary Pharmacogenomic Review Group (IPRG), with contributions of expertise from different FDA Centers, such as the Center for Drug Evaluation and Research (CDER), the Center for Biologicals Evaluation and Research, the Center for Devices and Radiological Health, and the National Center for Toxicological Research, as well as across clinical divisions and from nonclinical toxicology reviewers in CDER.

The IPRG Biomarker Qualification Review Team evaluates study protocols and reviews study results for the qualification of novel biomarkers of drug safety, using appropriate preclinical, clinical, and statistical considerations. The team then develops recommendations and guidance for the submission of biomarker data, assessing the original biomarker context proposal through voluntary data submission, and then evaluates the qualification study protocol together with the sponsor to reach a consensus protocol. Finally, this team reviews qualification study results and drafts a recommendation for the clinical divisions regarding the approval or rejection of the qualification submission.

THE ROLE OF THE CRITICAL PATH INITIATIVE AND THE BIOMARKERS CONSORTIUM IN BIOMARKERS DEVELOPMENT

It is evident that new investigational paradigms in drug development must be advanced to facilitate both discovery and clinical development, without sacrificing basic regulatory standards of safety and efficacy [33]. There are, however, obstacles to be overcome. Although biomarkers and surrogate endpoints have the potential to bring promising science to the clinic more expeditiously, there is as yet little agreement on the criteria for validating these new entities. The biomarker validation process itself is time-consuming and expensive. Intellectual property issues may also hamper validation. Perhaps the biggest hurdle is the need for stakeholders to agree that clinical investigation is not a perfect science, that uncertainty always has and always will remain at the end of the development process (particularly regarding safety), and that the use of biomarkers and surrogates of efficacy need not necessarily amplify that uncertainty.

Cognizant of these challenges, there is evidence that the stakeholders in the pharmaceutical enterprise (health care providers, regulatory authorities, industry, and payers) recognize the need for a shift in the approach to drug development [34]. Indeed, the U.S. FDA has put forward a Critical Path Initiative [35] that identifies a choice between the status quo, “stagnation,” and a new path, “innovation”, and describes critical path research as being “directed toward improving the product development process itself by establishing new evaluation tools”.

The many challenges of biomarker research and development that were clearly articulated by the FDA Critical Path Initiative[35] and their opportunities list [36] and The Biomarkers Definitions Working Group [13], has seen the development of several consortia in recent years [33].

These include the C-Path Institute (www.c-path.org/), Predictive Safety Testing Consortium (PSTC) the European Innovative Medicines Initiative (ec.europa.eu/research/health/imi/member-states-group_en.html), and The Biomarkers Consortium, which is a public-private partnership managed by the Foundation for the National Institutes of Health (FNIH; <http://www.FNIH.org>). The Biomarkers Consortium endeavors to identify, develop, and qualify biological markers (biomarkers) to support new drug development, preventive medicine, and medical diagnostics. The Biomarkers Consortium is a major public-private biomedical research partnership with broad participation from stakeholders across the health

enterprise, including government, industry, academia and patient advocacy and other non-profit private sector organizations.

CHALLENGES IN OSTEOARTHRITIS

Many hurdles exist within OA research and development that pertain to biomarker validation and qualification. The guidance and current gold standard for measuring clinical efficacy in disease modifying therapy development in OA is radiographic joint space narrowing (JSN) [37]. From joint space narrowing outcomes the health, integrity and thickness of hyaline articular cartilage are inferred [38, 39]. This guidance describes a process for drug approval for specific indications in OA, including treatment of symptoms, delays in structural progression and even discusses prevention of OA. The JSN measure is currently recommended by both the FDA and European Agency for the Evaluation of Medicinal Products (EMA) guidance documents as the imaging endpoint for clinical trials of disease-modifying OA drugs (DMOADs). At present, an alteration in structural progression would likely be determined by plain radiography, but it is possible that newer technologies may be approved including biochemical markers, MRI or even ultrasound, once appropriately validated. The FDA guidance is currently under review with efforts from an OARSI led initiative.

If we choose the current recommended endpoint, namely JSN, we would require hundreds of subjects, followed for at least 2–3 years, to demonstrate a significant incremental benefit of a novel therapy over and above that provided by currently available therapies. The direct costs of conducting such trials and the costs resulting from the overall duration of the therapeutic development and regulatory review process has dampened enthusiasm for development of therapeutic agents in this area and, in some instances, has rendered advancement of novel treatments prohibitively expensive. On the other hand, if other, more efficient means of establishing the benefit of new drugs exist, the promise of timely access to new therapies remains. There is, therefore, potentially tremendous value to public health in accelerating the discovery and development processes for OA therapeutics through smaller, shorter studies, using validated endpoints other than radiographic JSN. The use, in part, of clinical trial evidence based on biomarker and surrogate endpoint effects (in lieu of morbidity endpoints such as joint replacement) has the potential to revolutionize the drug development process and to thereby enhance the armamentarium of safe and effective therapeutics.

This accelerated path to new therapies in OA needs to be balanced by global concerns. Unlike other diseases where surrogate endpoints exist, OA does not have a mortality endpoint but rather affects a person's quality of life. Therefore, the 'clinical endpoint' is harder to establish. Furthermore, improvement in quality of life over the long interval of time that persons with this chronic disease receive therapy, can be easily dampened by toxicity, that can be fatal [40]. Thus the need for therapeutic advance needs to be balanced by not only demonstrating early efficacy but also ensuring sustainability of the effect and adequate safety.

Another challenge with the radiograph as the current reference standard is that it creates an imperfect reference for comparison with other methodologies for the purposes of validation [41]. Progression in joint space width (JSW) loss also reflects OA changes in joint tissues other than articular cartilage, particularly extrusion and degenerative changes of the menisci associated with OA development and progression [42]. If a purported therapeutic targeted synovium or bone marrow lesions directly, ascertaining its therapeutic benefit by the measurement of JSW may not be appropriate.

Measurement errors related to the variability in knee positioning requires considerable effort for the standardization of radiographic protocols, including the use of fluoroscopy, which has limitations including a lesser availability of the equipment, and greater radiation exposure to patients. In addition the relation between radiographic features and symptoms and other aspects of clinical outcome including joint replacement is not strong and frequently heterogeneous between studies [43]. Thus these limitations cast considerable doubt on our ability to use JSN as a single measure of efficacy in a clinical trial and allude to greater potential in other markers or alternatively a combination of measures.

Another challenge is that the current approval of potential therapies in OA requires that this structural alteration be linked to some clinical benefit either at the time when the structure was measured or at a later time-point. With this concept in mind it is obviously important that improvements in OA structural features are ascertained that are more likely linked to the clinical symptoms experienced by patients, or alternatively can serve as a surrogate for a clinically meaningful outcome. Similar to hypertension and osteoporosis, OA is a clinically silent disease for a long time. This extended lead time prior to clinical diagnosis is an opportunity for biomarkers to identify and link the early asymptomatic stages with the late stage classical radiographic indices of disease.

Currently there is little consensus on what constitutes a meaningful clinical endpoint for OA structure modifying trials. Some suggest that the development of symptomatic radiographic OA should suffice whereas others are developing definitions for what would constitute a virtual total joint replacement [44]. The lack of clear consensus creates an enormous challenge with regards to defining and validating efficacy biomarkers *let alone* the development of surrogate endpoints. Although the use of surrogate outcomes in clinical trials reduces sample size requirements and trial duration, they can only be justified if there is strong evidence that therapeutic targeting of the surrogate will translate into a beneficial patient outcome [22].

Additional challenges in biomarker validation in OA include:

1. Unlike cancer biomarkers for a specific cancer, OA is complicated by remarkable heterogeneity. This complexity sees different patterns of onset and clinical presentation, different patterns of joint involvement (hand vs. hip vs. knee etc), different patterns of compartment involvement within the same joint, different magnitude of synovial tissue involvement within the joint (e.g. variations in extent of meniscal, cartilage and bone involvement within subjects), and marked variations in the rate of disease progression. This adds to the considerable challenge both in determining the appropriate clinical outcome but also delineating efficacy biomarkers.
2. The current lack of a clear and reliably consistent disease modifying therapy tested using current biomarkers of efficacy does not permit comparative testing of novel biomarkers that encapsulate assessing the efficacy of an intervention. However in the interim, in the absence of effective treatments, validation studies can be performed in longitudinal observational studies and randomized controlled trials that have failed to demonstrate a treatment effect [45]. This is based upon the assumption they have appropriate specimens for biochemical markers, images for imaging biomarkers and adequate clinical outcomes data collected that will facilitate this purpose.
3. There has been no clear consensus of a pathway for biomarker validation and this uncertainty has stalled many development programs.

One recent step in the right direction was a manuscript that proposed the BIPED biomarker classification to give researchers a tool with which to classify biomarkers according to their intended purpose [46]. This classification system was ultimately developed to create a framework for depicting the wide range of biochemical markers and their respective purposes. This work was constructed with a focus on biochemical markers and a mandate for all biochemical marker uses. It touched upon validation but did not propose methods for validation or qualification. Thus whilst assisting in creating a framework it did not propose methods for particular biomarkers to be validated or qualified as suiting specific purposes such as Burden of Disease, Investigative, Prognostic, Efficacy of Intervention and Diagnostic purposes. The following section describes a suggested pathway that will address these concerns for OA.

PROPOSITION FOR OSTEOARTHRITIS

The lack of a clear consensus of a pathway for biomarker validation in OA can be overcome. What is described here is a suggested solution to this challenge. There is sufficient interest in OA biomarkers to warrant establishing a special interest group for OA within the biomarkers consortium. The Biomarkers Consortium (www.biomarkersconsortium.org) is dedicated to discovering new markers that detect and monitor disease and assess response to therapeutic agents [33, 47]. The first critical step for OA biomarker development and validation is to establish a process from consensus about what is required both in terms of process and infrastructure.

One likely step that warrants further consideration by this group is the determination of the clinical outcome of relevance in OA. Until there is consensus on this it will be challenging to validate biomarkers for intermediary mechanisms or disease processes.

Some additional items that warrant attention include (a) defining OA patient needs, (b) recommending an analytic paradigm for the translation of biomarker assays toward clinical validation and use in clinical trials and clinical care, (c) discussing ways biomarkers from different companies might be used in combination, (d) sharing of information regarding promising biomarkers developed in OA programs, and (e) becoming more familiar with platforms and techniques being developed in the private sector.

The establishment of a working group that would include biomarker experts from the academia, NIH, FDA, and industry could bring new solutions to OA biomarker development and accelerate implementation for OA therapeutics.

The immediate focus of this group could be reaching consensus on standardized methods for biomarker validation [12] and qualification [10] in OA and then with shared purpose pursuing the validation of specific biomarkers that could predict therapeutic response in the plethora of rich resources already available to us from well developed observational and clinical trial datasets. Where necessary development of de-novo datasets may be necessary. An alternate or complementary approach for consensus building on the merits of prospective OA biomarkers would be via application of the OMERACT filter [22] and/or hierarchically ranking the evidence status of biomarkers [48].

In order to facilitate this, an *infrastructure* for OA biomarker validation needs to be created. There are a number of existent models in other diseases of which the Early Cancer Detection Network provides a clear structure that could be replicated in our field of OA (<http://edrn.jpl.nasa.gov/>). An overview of the organizational structure is depicted in Fig. (1).

The components and projects that comprise the Early Detection Research Network include:

- Biomarker development laboratories: Responsible for the development and characterization of new biomarkers or the refinement of existing biomarkers.
- Biomarker Reference Laboratories: These serve as a Network resource for clinical and laboratory validation of biomarkers, which includes technological development, quality control, refinement, and high throughput.
- Clinical Epidemiology and Validation Centers: These conduct clinical and epidemiological research regarding the clinical application of biomarkers.
- Data Management and Coordinating Center: Coordinates the EDRN research activities, providing logistic support, and conducting statistical and computational research for data analysis, analyzing data for validation. Also responsible for EDRN common database development.
- Informatics Center: The Informatics Center supports EDRN's efforts through the development of software systems for information management. The center is at the NASA Jet Propulsion Laboratory at the California Institute of Technology.

For successful biomarker validation, there is a great need for good specimen collections as proposed and developed by the NCI and is facilitated through the ECRN [49]. As the technology is fast changing, sample storage and processing play a critical role in determining the suitability of the specimens for various technologies. Whenever possible, pristine samples should be prioritized for discovery efforts. Open interaction among steering committees of large trials and large cohort studies should be encouraged for the free exchange of ideas and specimens for biomarker validation. This will ensure proper use of specimens across the research community. By establishing networks of cooperative human tissue banks or resources, existing resources could be made available for discovery rather than through single-tissue banks. Additional resources should be established to catalogue tissue specimen collections acting as "tissue collector" with clearly developed standard operating procedures and common data elements.

Systematic monitoring of specimen quality over time should be established with measures to randomly test the quality of samples or other checks for maintaining the integrity of the specimen bank. Academic laboratories should be encouraged to adapt industrial standards for compliance of good laboratory practices and necessary quality control measures. Authorship and ownership issues should be clearly defined by these repositories to avoid future conflicts related to the use of specimens. The NCI has developed the National Biospecimen Network to address these issues [50, 51].

Consideration of prior experiences and processes will be invaluable for those pursuing these aims in OA. Not unlike rheumatoid arthritis where measures of disease activity are often a composite of a number of measures it is very likely that in OA we may have to use a composite of a number of different parameters to provide a measure that will adequately relate to the ultimate clinical outcome (whatever that may be). Current efforts using a composite of pain, function and structure in delineating the virtual joint replacement is one example of this methodology [44]. Often the measurement performance of a single biomarker is inadequate with regards isolating for example false positive and negative response to therapy and this misclassification may be improved by combining biomarkers [52, 53].

These biomarker validation projects are a shared interest of all in the OA field. The Biomarker Consortium is precompetitive with respect to traditional pharmaceutical or biotechnology research and development and may overcome the current "silo" approach in pursuit of the "pet" biomarker. The shared results, shared costs and collaborative flow of information for both for-profit and non profit parties can overcome key impediments to

biomarker qualification. All participants can benefit from being able to gain access to complex and innovative technology, test proprietary compounds using the data from biomarker research, and diminish financial risk by sharing the costs of these efforts between multiple funders. This is obviously not without challenges including satisfying the expectations of a diverse group of stakeholders, identification of funding for projects at an early stage of development, and harnessing the intellectual and resource potential of a disparate field. Many within our field will claim this path to disease modification is hopeless and similarly the rigorous pursuit of biomarker validation pointless when we have no perfect reference standard. However these challenges are not as great as those faced by our patients, who warrant collaborative pursuit of therapeutic advance for a disease where we need safe and effective therapeutic interventions. Ultimately they will be the major beneficiaries from the new insights provided into disease risk, characterization and treatment.

CONCLUSION

Improved knowledge of the pathogenesis of OA and of its molecular and anatomic pathology and the wealth of information relating biomarkers of disease with clinical outcomes will permit a better means for the assessment of the effects of new therapeutic interventions in OA. It is evident from conversations within our field between industry, government regulators, and academia that there is a shared recognition of the need for both the development and application of new and existing biomarkers in therapeutic development. To get to this point we need a process for biomarker validation and qualification in OA. This is a path other disease areas have taken and there are experiences, processes and infrastructure mechanisms in existence that we can build upon.

Acknowledgments

We greatly appreciate the thoughtful comments of Jeffrey N. Siegel, M.D. Team Leader of FDA/CDER/OND/ODE II/DAARP [Food & Drug Administration/Center for Drug Evaluation and Research/Office of New Drugs/Office of Drug Evaluation II/Division of Anesthesia, Analgesia and Rheumatology Products]. The views in this manuscript do not necessarily reflect those of the Food and Drug Administration.

The corresponding author had final responsibility for the decision to submit for publication. Dr Hunter receives research or institutional support from AstraZeneca, DonJoy, Lilly, Merck, NIH, Pfizer, Stryker and Wyeth.

References

1. Kluff C. Principles of use of surrogate markers and endpoints. *Maturitas*. 2004; 47(4):293–8. [PubMed: 15063482]
2. Wagner JA. Overview of biomarkers and surrogate endpoints in drug development. *Disease Markers*. 2002; 18(2):41–6. [PubMed: 12364809]
3. Cummings J, Ward TH, Greystoke A, Ranson M, Dive C. Biomarker method validation in anticancer drug development [Review] [103 refs]. *Br J Pharmacol*. 2008; 153(4):646–56. [PubMed: 17876307]
4. Richter WS. Imaging biomarkers as surrogate endpoints for drug development. *Eur J Nucl Med Mol Imag*. 2006; 33(Suppl 1):6–10.
5. DiMasi JA, Hansen RW, Grabowski HG. The price of innovation: new estimates of drug development costs. *J Health Economics*. 2003; 22(2):151–85.
6. Berndt, E.; Gottschalk, A.; Strobeck, M. Opportunities for improving the drug development process: results from a survey of industry and FDA. MIT-FDA-Industry White Paper. 2006. <http://web.mit.edu/cbi/docs/berndt-et-al6-3-05pdf>[cited 2009 Mar 17]; Available from: URL: <http://web.mit.edu/cbi/docs/berndt-et-al6-3-05.pdf>
7. Bingham CO III, Buckland-Wright JC, Garner P, et al. Risedronate decreases biochemical markers of cartilage degradation but does not decrease symptoms or slow radiographic progression in patients with medial compartment osteoarthritis of the knee: results of the two-year multinational

- knee osteoarthritis structural arthritis study. *Arthritis Rheumatism*. 2006; 54(11):3494–507. [PubMed: 17075851]
8. Temple R. Are surrogate markers adequate to assess cardiovascular disease drugs? *JAMA*. 1999; 282(8):790–5. [PubMed: 10463719]
 9. NIH Definitions Working Group. Biomarkers and surrogate endpoints in clinical research: definitions and conceptual model. In: Downing, G., editor. *Biomarkers and Surrogate Endpoints: Clinical Research and Applications*. Amsterdam: Elsevier; 2000. p. 1-9.
 10. Goodsaid FM, Frueh FW, Mattes W. Strategic paths for biomarker qualification. *Toxicology*. 2008; 245(3):219–23. [PubMed: 18280028]
 11. NIH-FDA Conference: Biomarkers and Surrogate Endpoints: Advancing Clinical Research and Applications. Abstracts. *Disease Markers*. 1998; 14(4):187–334. [PubMed: 10505081]
 12. Lesko LJ, Atkinson AJ Jr. Use of biomarkers and surrogate endpoints in drug development and regulatory decision making: criteria, validation, strategies. [Review] [61 refs]. *Annu Rev Pharmacol Toxicol*. 2001; 41:347–66. [PubMed: 11264461]
 13. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Therapeutics*. 2001; 69(3):89–95.
 14. Fleming TR. Surrogate endpoints and FDA's accelerated approval process. *Health Affairs*. 2005; 24(1):67–78. [PubMed: 15647217]
 15. Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? *Ann Intern Med*. 1996; 125(7):605–13. [PubMed: 8815760]
 16. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. *Stat Med*. 1989; 8(4):431–40. [PubMed: 2727467]
 17. US Food and Drug Administration. Guidance for industry - pharmacogenomic data submissions. 2005. <http://www.fda.gov/cder/guidance/6400fnlpdf>[cited 2009 Mar 20]; Available from: URL: www.fda.gov/cder/guidance/6400fnl.pdf
 18. Lee JW, Devanarayan V, Barrett YC, et al. Fit-for-purpose method development and validation for successful biomarker measurement. *Pharm Res*. 2006; 23(2):312–28. [PubMed: 16397743]
 19. Wagner JA, Williams SA, Webster CJ. Biomarkers and surrogate end points for fit-for-purpose development and regulatory evaluation of new drugs. *Clin Pharmacol Therapeutics*. 2007; 81(1):104–7.
 20. Goodsaid F, Frueh F. Biomarker qualification pilot process at the US Food and Drug Administration. *AAPS J*. 2007; 9(1):E105–8. [PubMed: 17408233]
 21. Rolan P. The contribution of clinical pharmacology surrogates and models to drug development--a critical appraisal. *Br J Clin Pharmacol*. 1997; 44(3):219–25. [PubMed: 9296315]
 22. Lassere M. A users guide to measurement in medicine. *Osteoarthritis Cartilage*. 2006; 14(Suppl 1):10–4.
 23. Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT filter for Outcome Measures in Rheumatology. *J Rheumatol*. 1998; 25(2):198–9. [PubMed: 9489805]
 24. Katz R. Biomarkers and surrogate markers: an FDA perspective. *NeuroRx*. 2004; 1(2):189–95. [PubMed: 15717019]
 25. Lassere MN. The Biomarker-Surrogacy Evaluation Schema: a review of the biomarker-surrogate literature and a proposal for a criterion-based, quantitative, multidimensional hierarchical levels of evidence schema for evaluating the status of biomarkers as surrogate endpoints. *Stat Methods Med Res*. 2008; 17(3):303–40. [PubMed: 17925313]
 26. Li Z, Chines AA, Meredith MP. Statistical validation of surrogate endpoints: is bone density a valid surrogate for fracture? *J Musculoskeletal Neuronal Interact*. 2004; 4(1):64–74.
 27. Hughes MD, DeGruttola V, Welles SL. Evaluating surrogate markers [Review] [22 refs]. *J Acquired Immune Deficiency Syndr Hum Retrovirol*. 1995; 10(Suppl 2):S1–S8.
 28. Boissel JP, Collet JP, Moleur P, Haugh M. Surrogate endpoints: a basis for a rational approach [Review] [19 refs]. *Eur J Clin Pharmacol*. 1992; 43(3):235–44. [PubMed: 1425885]
 29. Consensus report of the Working Group on: "Molecular and biochemical markers of Alzheimer's disease". The Ronald and Nancy Reagan Research Institute of the Alzheimer's Association and the

- national institute on aging working group. [erratum appears in *Neurobiol Aging* 1998; 19(3): 285]. *Neurobiol Aging*. 1998; 19(2):109–16. [PubMed: 9558143]
30. De Gruttola VG, Clax P, DeMets DL, et al. Considerations in the evaluation of surrogate endpoints in clinical trials. summary of a National Institutes of Health workshop. *Control Clin Trials*. 2001; 22(5):485–502. [PubMed: 11578783]
 31. Weir CJ, Walley RJ. Statistical evaluation of biomarkers as surrogate endpoints: a literature review. *Stat Med*. 2006; 25(2):183–203. [PubMed: 16252272]
 32. Goodsaid F, Frueh F. Process map proposal for the validation of genomic biomarkers. *Pharmacogenomics*. 2006; 7(5):773–82. [PubMed: 16886901]
 33. Altar CA. The Biomarkers Consortium: on the critical path of drug discovery. *Clin Pharmacol Therapeutics*. 2008; 83(2):361–4.
 34. Revkin JH, Shear CL, Pouleur HG, Ryder SW, Orloff DG. Biomarkers in the prevention and treatment of atherosclerosis: need, validation, and future. *Pharmacological Rev*. 2007; 59(1):40–53.
 35. FDA. Critical Path Initiative. 2009. <http://www.fda.gov/oc/initiatives/criticalpath/> [cited 2009 Mar 17]; Available from: URL: <http://www.fda.gov/oc/initiatives/criticalpath/>
 36. FDA. Critical Path Opportunities Report and List. 2006. http://www.fda.gov/oc/initiatives/criticalpath/reports/opp_reportpdf [cited 2009 Mar 17]; Available from: URL: http://www.fda.gov/oc/initiatives/criticalpath/reports/opp_report.pdf
 37. Food and Drug Administration. Guidance for Industry. Clinical Development Programs for Drugs, Devices, and Biological Products Intended for the Treatment of Osteoarthritis (OA). 1999. <http://www.fda.gov/Cber/gdlns/osteohm> [cited 2009 Mar 17]; Available from: URL: <http://www.fda.gov/Cber/gdlns/osteo.htm>
 38. Mazza SA, Brandt KD. Is knee radiography useful for studying the efficacy of a disease-modifying osteoarthritis drug in humans? *Rheum Dis Clin North Am*. 2003; 29(4):819–30. [PubMed: 14603585]
 39. Mazza SA, Brandt KD, Buckwalter KA, Lequesne M. Pitfalls in the accurate measurement of joint space narrowing in semiflexed, anteroposterior radiographic imaging of the knee. *Arthritis Rheum*. 2004; 50(8):2508–15. [PubMed: 15334464]
 40. Alpert JS. The Vioxx debacle. *Am J Med*. 2005; 118(3):203–4. [PubMed: 15745713]
 41. Guermazi A, Burstein D, Conaghan P, et al. Imaging in osteoarthritis. [Review] [183 refs]. *Rheum Dis Clin North Am*. 2008; 34(3):645–87. [PubMed: 18687277]
 42. Hunter DJ, Zhang YQ, Tu X, et al. Change in joint space width: hyaline articular cartilage loss or alteration in meniscus? *Arthritis Rheum*. 2006; 54(8):2488–95. [PubMed: 16868968]
 43. Hannan MT, Felson DT, Pincus T. Analysis of the discordance between radiographic changes and knee pain in osteoarthritis of the knee. *J Rheumatol*. 2000; 27(6):1513–7. [PubMed: 10852280]
 44. Gossec L, Hawker G, Davis AM, et al. OMERACT/OARSI initiative to define states of severity and indication for joint replacement in hip and knee osteoarthritis. *J Rheumatol*. 2007; 34(6):1432–5. [PubMed: 17552070]
 45. Mildvan D, Landay A, De GV, Machado SG, Kagan J. An approach to the validation of markers for use in AIDS clinical trials [Review] [184 refs]. *Clin Infect Dis*. 1997; 24(5):764–74. [PubMed: 9142767]
 46. Bauer DC, Hunter DJ, Abramson SB, et al. Classification of osteoarthritis biomarkers: a proposed approach. *Osteoarthritis Cartilage*. 2006; 14(8):723–7. [PubMed: 16733093]
 47. Zerhouni EA, Sanders CA, von Eschenbach AC. The Biomarkers Consortium: public and private sectors working in partnership to improve the public health. *Oncologist*. 2007; 12(3):250–2. [PubMed: 17405889]
 48. Lassere MN, Johnson KR, Boers M, et al. Definitions and validation criteria for biomarkers and surrogate endpoints: development and testing of a quantitative hierarchical levels of evidence schema. [24 refs]. *J Rheumatol*. 2007; 34(3):607–15. [PubMed: 17343307]
 49. Maruvada P, Srivastava S. Joint National Cancer Institute–Food and Drug Administration workshop on research strategies, study designs, and statistical approaches to biomarker validation for cancer diagnosis and detection. *Cancer Epidemiol Biomarkers Prevent*. 2006 Jun; 15(6):1078–82.

50. Birmingham K. An inauspicious start for the US National Biospecimen Network. *J Clin Investig.* 2004; 113(3):320. [PubMed: 14755325]
51. Hede K. NCI's National Biospecimen Network: too early or too late? *J Natl Cancer Inst.* 2005; 97(4):247–8. [PubMed: 15713954]
52. Feng Z, Yasui Y. Statistical considerations in combining biomarkers for disease classification. *Dis Markers.* 2004; 20(2):45–51. [PubMed: 15322313]
53. McIntosh MW, Pepe MS. Combining several screening tests: optimality of the risk score. *Biometrics.* 2002; 58(3):657–64. [PubMed: 12230001]

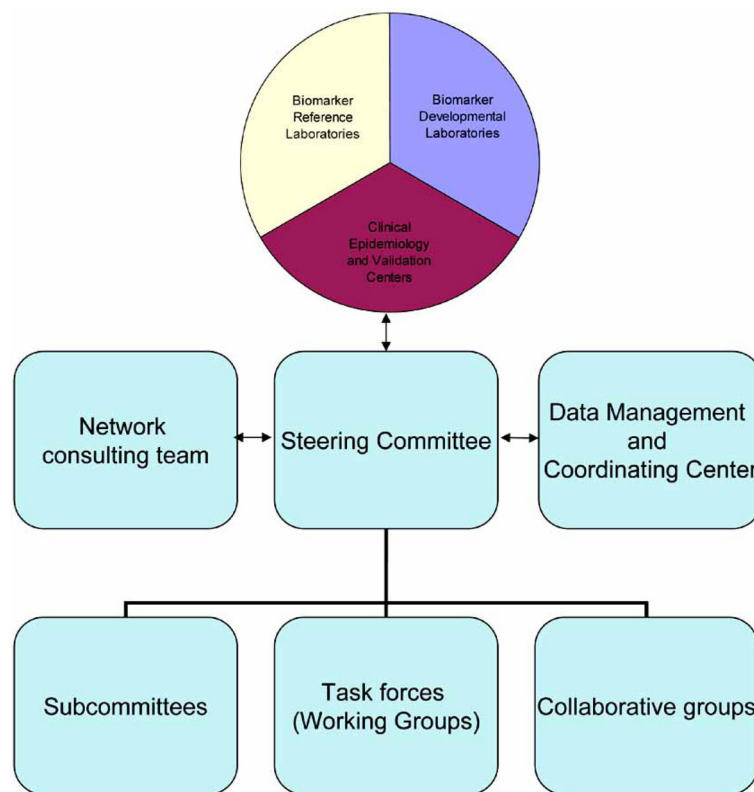


Fig. 1. Chart depicting the organization of EDRN. Modified from: http://edrn.jpl.nasa.gov/about-edrn/image-1.gif/image_view