Considerations for Use of Histopathology and Its Associated Methodologies to Support Biomarker Qualification

Guidance for Industry

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)

May 2016
Procedural

Considerations for Use of Histopathology and Its Associated Methodologies to Support Biomarker Qualification

Guidance for Industry

Additional copies are available from:

Office of Communications, Division of Drug Information
Center for Drug Evaluation and Research
Food and Drug Administration
10001 New Hampshire Ave., Hillandale Bldg., 4th Floor
Silver Spring, MD 20993-0002
Phone: 855-543-3784 or 301-796-3400; Fax: 301-431-6353
Email: druginfo@fda.hhs.gov
http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)

May 2016
Procedural

TABLE OF CONTENTS

Contains Nonbinding Recommendations

Considerations for Use of Histopathology and Its Associated Methodologies to Support Biomarker Qualification Guidance for Industry[1]

This guidance represents the current thinking of the Food and Drug Administration (FDA or Agency) on this topic. It does not establish any rights for any person and is not binding on FDA or the public. You can use an alternative approach if it satisfies the requirements of the applicable statutes and regulations. To discuss an alternative approach, contact the FDA office responsible for this guidance as listed on the title page.

I. INTRODUCTION

This guidance is intended to assist submitters of a **biomarker**[2] for qualification that conduct nonclinical studies for which histopathology is used as a reference or **truth standard** (Zhou et al. 2011). Scientifically rigorous evaluation of biomarker performance in relation to histopathologic changes is essential in these studies because they may provide direct evidence to support nonclinical biomarker qualification or supportive translational data to aid in the development and qualification of clinical biomarkers for a proposed context of use. This guidance discusses the issues that should be considered when generating histopathology data in nonclinical **biomarker qualification studies** and outlines the scientific standards recommended for biomarker characterization and qualification. The biomarker qualification process is described in a separate guidance, *Qualification Process for Drug Development Tools*.[3]

In general, FDA's guidance documents do not establish legally enforceable responsibilities. Instead, guidances describe the Agency's current thinking on a topic and should be viewed only as recommendations, unless specific regulatory or statutory requirements are cited. The use of the word *should* in Agency guidances means that something is suggested or recommended, but not required.

[1] This guidance has been prepared by the Office of New Drugs and the Office of Translational Sciences in the Center for Drug Evaluation and Research (CDER) at the Food and Drug Administration.
[2] Terms that appear in bold type upon first use are defined in the Glossary section of this guidance.
[3] See FDA guidance for industry and FDA staff on *Qualification Process for Drug Development Tools*, available at *http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm*. We update guidances periodically. To make sure you have the most recent version of a guidance, check the FDA Drugs guidance Web page at http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm.

II. BACKGROUND

The discovery, characterization, qualification, and use of biomarkers have been identified by the FDA Critical Path Initiative[4] as important means for improving the efficiency and success rate of drug[5] development.

A **biological marker** or biomarker is defined as a characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes, or biological responses to a therapeutic intervention (Biomarkers Definitions Working Group 2001).

The context of use is a comprehensive and clear statement that describes the manner of use, interpretation, and purpose of use of a biomarker in drug development. Details about the elements of context of use and categories of biomarkers can be found at the Biomarker Qualification Context of Use Web page (http://www.fda.gov/Drugs/DevelopmentApprovalProcess/DrugDevelopmentToolsQualificationProgram/ucm284620.htm). For nonclinical biomarkers, submitters should specify whether the intended use of the biomarker in nonclinical studies will be in normal, healthy animals or in animal models of disease.

When a biomarker is qualified, analytically valid measurements of it can be relied upon to have a specific and interpretable meaning (e.g., physiologic, toxicologic, pharmacologic, or clinical) in drug development and regulatory decision-making. Industry can then use the biomarker for its qualified context of use during premarketing drug development, and FDA reviewers can be confident about its qualified context of use without the need to reconfirm its applicability or utility. The qualification process includes a formal review to evaluate data supporting a biomarker for a particular context of use. The types of studies to be conducted in support of a biomarker qualification will depend upon the proposed context of use, which dictates the depth and extent of data. Accordingly, data supporting a nonclinical biomarker qualification should be reliable and repeatable. The biomarker also can be used for a purpose outside the qualified context of use in investigational new drug application (IND) submissions, subject to review and discussion with CDER, on a case-by-case basis.

The terms **exploratory study** and **confirmatory study** are used in this guidance similarly to their use in describing human efficacy studies — to refer to **exploratory nonclinical biomarker qualification studies** and **confirmatory nonclinical biomarker qualification studies**, respectively. Exploratory studies should have clear objectives. The results of the studies may provide initial information to guide further investigation. These studies usually occur early in the characterization of a biomarker and can use a flexible design that allows for changes in the study protocol and analyses in response to the accumulating data. The studies are often informative but would generally not be the sole source of data to support the proposed context of use of a biomarker. In contrast, a confirmatory study is well designed, well controlled, and

[4] See FDA's Critical Path Initiative, available at http://www.fda.gov/scienceresearch/specialtopics/criticalpathinitiative/default.htm.

[5] The term *drug* as used in this guidance refers to both human drugs and therapeutic biologics regulated by CDER unless otherwise specified.

hypothesis driven. Confirmatory studies are expected to provide substantive evidence to support the proposed context of use of a biomarker.

A qualification determination is a regulatory conclusion made by the FDA. **Qualification study** is a general term used in this guidance to refer to nonclinical studies having to do with the process of biomarker qualification. Data quality and integrity are important to support FDA's conclusion that the biomarker should be qualified for its specific context of use. A qualified biomarker is intended to support drug development and regulatory decisions. Thus, the qualification data should be reliable and repeatable.

Traditionally, histopathology has been used to identify morphologic changes associated with an in vivo diagnosis, evaluation of response to therapy, basic research, and nonclinical safety assessment. There is often a strong correlation between specific histopathologic findings and clinical signs, or some clinical chemistry parameters. Thus, histopathology is currently used in some biomarker qualification programs as a reference or truth standard to evaluate the temporal correlation of the biomarker with the evolution and reversibility of the morphologic changes. This guidance is intended to provide recommendations to investigators who conduct nonclinical biomarker qualification studies when histopathology is used as a reference or truth standard.

This guidance is not intended to address the use of histopathology in traditional drug development nonclinical safety assessment studies. Biomarker qualification studies are different from nonclinical safety assessment studies and call for a different scientific paradigm. Nonclinical safety assessment studies are conducted primarily for hazard characterization of a specific drug, and histopathologic assessment assists in identifying effects of the drug. In safety assessment studies, it is common practice for the pathologist to evaluate the histopathology knowing the identity of the treatment groups, the clinical signs, hematology, clinical chemistry, urinalysis, organ weight, and gross pathology data, and then to integrate this information in a unified, physiologically plausible format (Crissman et al. 2004).

In contrast, histopathology is used in nonclinical biomarker qualification studies to establish correlative and temporal relationships between the biomarker and morphologic changes. Examples include (1) establishing a possible quantitative relationship between biomarker level and severity of morphological change; (2) exploring the kinetics of release, clearance of the biomarker, and correlative histomorphology; (3) determining the range of normal variability of the biomarker and any associated morphologic findings; and (4) identifying potentially confounding factors for use of the biomarker. Particularly in confirmatory nonclinical biomarker qualification studies, histopathology can be used as a reference or truth standard. In these cases, the pathologist should function as an impartial interpreter, uninfluenced by other information, and unequivocally blinded to novel biomarker value and comparator value (e.g., standard clinical chemistry value) and preferably to assigned treatment group.

Although this guidance is not intended to apply to the conduct of nonclinical safety assessment studies, biomarker data, either exploratory or confirmatory, can be gathered from safety assessment studies to support the qualification of a biomarker. For example, data collected following "Best Practices," such as those outlined by Burkhardt et al. 2011, can be used as supportive data, regardless of whether the data were derived from safety assessment studies or

dedicated qualification studies. We recommend that biomarker developers consult with the FDA prior to initiating dual purpose safety assessment-qualification studies.

Qualification of a biomarker is a regulatory conclusion, and thus, the studies used to support qualification are expected to be scientifically valid, of high quality, and unbiased.

III. ESTABLISHING PERFORMANCE CHARACTERISTICS OF BIOMARKERS

In this section, we describe **biomarker sensitivity, biomarker specificity, analytical sensitivity,** and **analytical specificity (**See **Sensitivity** and **Specificity** in Glossary). **Diagnostic sensitivity** and **diagnostic specificity** of in vitro diagnostic tests are also defined in the Glossary (see **Sensitivity** and **Specificity** in Glossary) for clarity, but are separate from biomarker qualification.

A. Methodology for Detection of the Biomarker (Biomarker Assay)

Biomarker detection can be based on several different methodologies (e.g., biochemical measurements, physiologic organ function tests, or imaging of structural features ranging from molecular to anatomic levels). The method of biomarker detection provides the measurements that are the basis for comparison of the biomarker to the reference or truth standard of histopathology and, therefore, is critical to biomarker characterization. Thus, the detection system should be well characterized.[6]

For example, for a biochemical biomarker, the assay used to measure the biomarker should have scientifically rigorous analytical sensitivity and analytical specificity. An assay with poor sensitivity or one that is inhibited by other substances in the biological sample or matrix could lead to a false negative result by failing to detect initial changes in biomarker levels. An assay with poor specificity would not distinguish injury to a nontarget organ or the assay may cross-react with nonspecific substances in the specimen. Thus, the biomarker assay should reliably and reproducibly detect changes (increase or decrease) or absence of changes in biomarker levels. Interpretation of any detected change in biomarker levels should ultimately include an assessment of whether the change is biologically meaningful through comparison to the histopathology.

Considerations for biomarker detection based on biochemical measurements also apply to biomarker detection based on other methodologies such as imaging technologies (e.g., positron emission tomography, magnetic resonance imaging, radiology). Histopathology used as a reference or truth standard provides an independent means of evaluating the same variable that is assessed by the biomarker.

[6] See FDA's draft guidance for industry *Bioanalytical Method Validation* (Revision 1). In the *Federal Register* of September 13, 2013 (78 FR 56718), FDA published a notice announcing the availability of the revised draft guidance. When finalized, it will represent the Agency's current thinking on the topic.

B. Biological Performance of the Biomarker

1. Performance Characteristics

It is highly unlikely that any biomarker will be 100 percent specific or 100 percent sensitive. The proposed context of use for a biomarker determines the specific performance characteristics that can be used to support qualification. Studies intended to establish the sensitivity, specificity, and reproducibility of the biomarker should be relevant to the proposed context of use. A reference or truth standard such as histopathology provides an independent means of evaluating the same parameter that the biomarker is intended to assess.

2. Biomarker Sensitivity and Temporal Correlation

In contrast to analytical sensitivity, biomarker sensitivity is related to (1) the probability of identifying a morphologic effect, (2) the threshold of morphologic change that would cause a discernible change in the biomarker level, and (3) the time interval before or after a morphologic effect until a change in the biomarker level can be detected.

Biomarker sensitivity is the probability that a biomarker will indicate a specific morphologic change when the change is present. If the biomarker is 100 percent sensitive, it will never be negative in the presence of the specific morphologic change that it is designed to detect. The temporal relationship between the initial change in biomarker level and the onset of histopathologic change defines the biomarker sensitivity. Accurate characterization of this temporal relationship is important to optimizing the biomarker sensitivity.

The timing and frequency of biomarker measurement in qualification studies should allow for a clear interpretation of the temporal relationship of the biomarker with morphologic changes. The time frame and mechanism by which an agent (e.g., mechanical, chemical, natural disease) produces morphologic changes in tissue should be understood and considered when planning the time course of sampling and/or analysis in biomarker qualification studies.

Histopathologic evaluation using light microscopy with hematoxylin and eosin (H&E) staining is preferred in qualification studies, when this is the method that will be used in the context of use of the biomarker. For treatment-induced histopathologic changes in animals, the temporal relationship between changes in biomarker levels and detectable morphologic changes, as observed by light microscopy and H&E staining, usually can be described by one of the following statements, as indicated by appropriate supportive data:

- The initial change in biomarker levels precedes the onset of morphologic changes.
- The initial change in biomarker levels is approximately concurrent with the onset of morphologic changes.
- The initial change in biomarker levels occurs only after the development of morphologic change.

Similarly, when there is a relationship between morphologic and biomarker changes during reversal or recovery, morphologic changes may be preceded by, concurrent with, or followed by changes in the biomarker levels. Sometimes recovery from morphologic changes may be unrelated to changes in the biomarker levels.

There are several potential causes for false negative biomarker results (when a morphologic effect is not accompanied by an expected change in biomarker). These include (1) low analytical sensitivity, (2) interference of other proteins or substances with the assay, (3) inconsistent expression of the biomarker under different physiologic or pathologic conditions, (4) a prolonged latency in biomarker change relative to the morphologic effect, (5) the morphologic effect being caused by a mechanism other than the one measured by the biomarker, (6) mishandling of a specimen, or (7) inadequate sampling of a specimen.

Histopathologic sensitivity depends on the histopathologic methods of examination. The use of light microscopy for the initial evaluation of biomarker sensitivity is preferable. Special methods (e.g., electron microscopy, immunohistochemistry) may be useful to investigate whether small or early changes in biomarker levels are associated with histopathologic changes too subtle to be detected by light microscopy with H&E stain.

Biomarker qualification studies should evaluate the quantitative correlation between the magnitude of biomarker change and the degree of morphologic change. For these studies, it is critical to obtain a final biomarker measurement as close to the tissue collection as possible (at least on the same day) to avoid a temporal disconnect in these two parameters.

3. Biomarker Specificity

The specificity of a biomarker is the probability of correctly determining that no morphologic change has occurred in the tissue of interest. A biomarker that has 100 percent specificity will result in no false positives. Specificity also can be viewed as a very low or very stable baseline value when there is no morphologic change. A highly specific biomarker will not have a signal originating in unaffected or non-target tissues. Appropriate supportive data helps to establish specificity of a biomarker.

The specificity of a biomarker candidate can be challenging to evaluate, particularly when it is known that multiple isoforms exist with different tissue distributions. For example, kidney injury molecule-1has isoforms specific to liver and kidney (Bailly et al. 2002); troponins have isoforms specific to cardiac and skeletal muscle (reviewed in Wei and Jin 2011). In such cases, it is important to examine histologic and biomarker isoform expression in multiple tissue types to assure that the biomarker candidate has specificity for the tissue of interest.

There are several causes of a false positive biomarker signal. Examples include low analytical specificity, the release of the biomarker from a nontarget tissue, or expression of the biomarker under both physiologic and pathologic conditions.

As noted previously, it is highly unlikely that any biomarker will be 100 percent specific or 100 percent sensitive. The proposed context of use of the biomarker will define the appropriate

balance between biomarker sensitivity and specificity. Sometimes a single biomarker does not fulfill a particular context of use, and therefore the use of multiple biomarkers may be considered. We recommend that the use of multiple biomarkers be discussed with the Agency during design of the confirmatory studies for biomarker qualification to ensure that all Agency concerns are addressed.

4. Reversibility/Resolution

Studies designed to investigate the recovery from or the cessation of a physiologic or pathologic process should characterize the temporal relationship between the reversal of biomarker changes and the cessation/recovery of the histopathologic changes. Note, however, that the reversal of the biomarker changes may not correlate with the cessation or recovery of the histopathologic changes. For example, in the context of drug-induced injury, reversal of the biomarker levels may be caused by the cessation of the biomarker release from the site of injury and/or the subsequent degradation or elimination from the body, without cessation/recovery of the histopathologic changes. Thus, an adequate interpretation of the biomarker and histopathologic changes is complex and all data should be considered.

5. Alternative Testing Models

An alternative to the use of healthy animals for nonclinical biomarker qualification is the use of animal models of human disease. These animal models may be useful for evaluating the change in a biomarker candidate in response to a pharmacologic, physiologic, or physical intervention.

IV. SPECIFIC ASPECTS OF HISTOPATHOLOGIC METHODS IN CONFIRMATORY BIOMARKER STUDIES

The confirmatory biomarker qualification studies generally will be conducted after many of the early questions on the proposed context of use and the assays have been refined. The focus of the histopathology, such as the tissue of interest, the specific lesions, and important features of lesions, will have been identified before planning and initiation of confirmatory studies. The topics discussed in this section are critical to confirmatory studies and should be considered and addressed to provide convincing study results.

A. General Planning for Confirmatory Studies

The first step generally entails a systematic review of the literature (Mignini and Khan 2006; Piper et al. 1996; Pound et al. 2004; Roberts et al. 2002) conducted in a way that avoids bias in the search. A literature review can be used both to plan the qualification studies and to support a qualification effort.

An important factor in planning confirmatory studies is a prospectively written protocol that discusses all steps leading to and including the evaluation of the histopathology. All aspects of the necropsy, including timing, handling of the specimen and tissues, fixation, sample storage, transportation (if relevant), and slide preparation and evaluation, should be described clearly.

Some of the important points that should be considered are discussed in the following subsections.

B. Slide Examination Procedures to Minimize Bias in Confirmatory Studies

A pathologist should be free to conduct a blinded assessment of outcomes for any study, especially if that is the standard practice for the institution (Holland and Holland, 2011a; Holland and Holland, 2011b). Recommendations outlined by Burkhardt et al. provide criteria for when it may be appropriate to use blinded evaluations, taking into account the treatment groups and time course to characterize changes, devising quantifying and qualifying scoring systems, and determining thresholds for background changes (Burkhardt et al. 2011).

For exploratory studies used to support confirmatory studies, a tiered analysis approach may be appropriate, using procedures as proposed by the Society for Toxicologic Pathology (Crissman et al. 2004). The first level in this tiered approach can be an unblinded comparison of treated and control specimens to identify subtle findings (such as the increase in incidence or severity of spontaneous findings) and to develop scoring criteria. Following this first tier analysis, the pathologist might perform a targeted, blinded evaluation from any or all groups as appropriate. This allows for identification of subtle, treatment-related findings that can be consistently differentiated from those that occur in controls. A **pathology peer review**, with targeted blinding by a second pathologist who is naïve to the study, may also help minimize bias (Morton et al. 2010). Lastly, enlisting a **pathology working group** may be valuable.

The goal of a confirmatory study is to test whether the biomarker accurately reflects the histopathology to support the proposed context of use. Subjective judgment can influence any step of slide evaluation, defeating the purpose of the study. The risk of bias is greater when the histopathologic changes under evaluation are subtle or sparse (the level of tissue change of most interest in many biomarker qualification studies). Slide examination procedures should be defined prospectively in the confirmatory study protocol, including steps to minimize bias. In confirmatory studies, we strongly recommend that tissue sections be evaluated with the reader blinded to treatment condition, sampling times, novel biomarker results, and any comparator biomarker results.

Slides from concurrent control group animals can be used initially to establish background morphology and aid in distinguishing subtle histopathologic changes from normal variation. This assessment of normal variations in background histomorphology is important for identifying thresholds to distinguish treatment-related lesions and to understand whether certain background histomorphology is associated with altered biomarker values. Extra slides from control group animals can be prepared in sufficient number to allow assessment of normal variations and background lesions. When acquiring additional sections from the planned tissue samples is not feasible, additional control group animals can be included in the nonclinical studies. For statistical purposes, it is preferable that the data derived from these extra slides not be included in the final dataset of the qualification study. The overall peer review assessment plan, if any, should be stipulated in the scientific plan and reflected in the protocol, along with a plan for resolving disagreements among readers/reviewers.

C. Timing of Sample Collection

The proposed context of use of a biomarker and the specific goal of a given study will determine the timing of tissue sample collection. It is important to determine the variability of biomarker measurements and the likely reasons for the variability. Timing of sample collection may be an important factor in variability. Thus, specific sampling times (study day and hour post-procedure, if applicable) should be selected carefully, and the rationale should be included in the protocol to facilitate interpretation of data and reproducibility.

D. Controls

The objective of a confirmatory study is to prospectively evaluate the performance of the biomarker relative to histomorphologic changes. The confirmatory study should be designed with sufficient rigor to permit a valid comparison of treatment and control groups and permit unbiased assessment of histologic material. To reduce the potential for analytical and biological false negative and false positive results, concurrent positive and negative control groups should be used to control for environmental factors, enable determination/comparison of biomarker sensitivity/specificity, and ensure reagent/method adequacy.

The identification of concurrent control groups and the rationale for their use should be described in the protocol. Adequate **historical (external) control** data can be used to identify potential analytical and biological outlier values in datasets; however, applying the historical data in this regard should be determined a priori and prospectively detailed in the protocol. Historical control data should be current and contain only studies for which similar control test agents and/or conditions were used (Keenan et al. 2009).

E. Fixation

Fixation procedures should be described in the study protocol. Variations in the time to fixation and duration of fixation may create variability in the data that obscures a true biomarker-histomorphologic relationship. Delays and variations in fixation of tissue should be avoided, and deviations from the protocol should be noted. To avoid experimental bias, animals should be randomized for scheduled necropsy.

F. Number of Sections and Sampling Location (at Necropsy)

The anatomic sites from which tissue is collected and the number of samples collected may affect results and their interpretation. Based on the stated goals, a rationale for the number of tissue sections to be collected and the locations from which they will be taken should be provided in the protocol. When deciding the number of sections to be collected and analyzed, a submitter should consider that there is a greater likelihood of missing subtle or sparse lesions if an insufficient number of samples are taken. A description of each sampling location should be sufficient to allow reproducibility across laboratories. The tissue/organ, lobe, and regional position within tissue (e.g., left, right, anterior, posterior, superior, inferior) from which samples will be obtained should be described in the protocol. If serial or step tissue sections will be

obtained, the procedures for identifying and reporting the precise location of lesions within an organ/tissue should be described in the protocol.

G. Staining

Assessment of histologic slides under the conditions of the proposed context of use is generally preferable. For example, most safety assessment studies use light microscopy for H&E-stained sections. The rationale should be provided for each stain used. If staining methods are performed in the laboratory according to previously established procedures, this information can be incorporated into the protocol as an appendix or it can be referenced. For nonstandard stains, the methods and scoring systems should be well documented or referenced in the protocol. We encourage using automated or standardized staining techniques for all samples, including positive and negative controls.

H. Special Methodology

In some cases (e.g., to confirm the morphologic process, cell type, or location), there may be a scientific need for more specific information or additional supportive data to achieve the study objectives. In these cases, special methodologies may be necessary to assess the morphology of particular targets to support a stated hypothesis. These methods include, but are not limited to, the following:

- stereology
- histochemistry
- immunohistochemistry
- in situ hybridization
- electron microscopy

The rationale for using a special methodology should be provided, and the methodology used should be well documented or referenced in the protocol.

V. OTHER CONSIDERATIONS FOR HISTOPATHOLOGIC EVALUATION

Other factors may affect the outcome of the slide evaluation. This section discusses certain factors that should be considered and addressed in the study protocol.

A. Digital Pathology and Slide Sharing

Digital slides should be shared and examined according to the established guidelines of recognized societies of toxicologic pathology (Tuomari et al. 2007) and the Digital Pathology Association (DPA) (see https://digitalpathologyassociation.org/).

B. Lexicons

Lexicons specify the predetermined criteria that classify types and degrees of change. Illustrations of the terminology are of great benefit in reducing variability.

For a well-defined morphologic condition, combining associated lesions for data entry is appropriate (e.g., chronic progressive nephropathy of rats). However, combining terms for ease of statistical analysis should not obscure individual characteristics appropriate for the unbiased analysis of the data and correlation with the biomarker.

C. Filtering

It is standard practice in diagnostic pathology to interpret certain lesions as background and not report them in results (filtering). Different pathologists can also use different thresholds for defining changes as background, thereby affecting the overall interpretation.

We recommend that data be generated to show how much, if at all, the background changes contribute to the normal variability of the biomarker (e.g., report all lesions, along with the associated biomarker values, independent of the filtering). An alternative is to comprehensively document the type and extent of lesions considered to be incidental background. The procedures for describing and documenting the background lesions should be described in the scientific plan or protocol.

D. Other Factors

Factors such as **diagnostic drift** and **chronological bias** are also important to consider, and procedures should be included in the protocol to control for these.

1. Diagnostic Drift

Diagnostic drift is defined as a gradual change in nomenclature or severity grading of lesions within a single study. It is a source of inconsistency that can negatively affect detection of treatment-related lesions/changes or the determination of no-effect levels (Crissman et al. 2004).

2. Chronological Bias

Chronological bias is defined as the evolutionary process of a grading system, whereby more specific and sensitive criteria for grade assignment are clarified, learned, and publicized. As pathologists gain experience, subtleties of the evolving system are applied to interpretation of tissue sections (Kondylis et al. 2003). This is also a source of inconsistency that can negatively affect detection of treatment-related lesions/changes or the determination of no-effect levels.

GLOSSARY

Biomarker (or biological marker): A characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes, or biological responses to a therapeutic intervention (Biomarkers Definitions Working Group 2001).

Biomarker qualification study or **qualification study:** A general term used in this guidance to refer to nonclinical studies specifically pertaining to the qualification of a biomarker.

Chronological bias: The evolutionary process of a grading system, whereby more specific and sensitive criteria for grade assignment are clarified, learned, and publicized (Kondylis et al. 2003).

Confirmatory nonclinical biomarker qualification study or **confirmatory study:** A well-designed and -controlled nonclinical study in which the hypotheses are stated in advance and evaluated. In such studies, the key hypothesis of interest is predefined and is the hypothesis that is tested when the study is complete. Confirmatory studies are intended to provide firm evidence in support of the proposed context of use of the biomarker and, therefore, adherence to the protocol is particularly important.

Diagnostic drift: A gradual change in nomenclature or severity grading of lesions within a single study. Diagnostic drift is a source of inconsistency that can negatively affect detection of treatment-related lesions/changes or the determination of no-effect levels (Crissman et al. 2004).

Exploratory nonclinical biomarker qualification study or **exploratory study:** A study that has clear objectives and provides information that may guide further investigation. In contrast to confirmatory nonclinical studies, an exploratory study may occur early in the characterization of a biomarker and can use a flexible design that allows for changes in the study protocol in response to the accumulating data. The analyses may involve exploring data and testing different hypotheses. Such studies cannot be the sole basis to support the proposed context of use of a biomarker, but can contribute to the total body of evidence. Exploratory studies are frequently (but not limited to) descriptive science.

Historical (external) control: A group of subjects, treated or untreated, from an earlier time than the study under consideration. Historical control data may serve to produce a range of normal rates of certain findings or changes (Keenan et al. 2009).

Lexicons: Consist of predetermined criteria used to classify types and degrees of change. A lexicon may be illustrated to clarify these predetermined criteria.

Pathology peer review: A secondary review with the objective of assisting the study pathologist to refine, verify, and improve the accuracy and quality of the final pathology assessments and interpretations.

Pathology working group: A group composed of the study pathologist, peer review pathologist, and at least one other pathologist. This group may be convened to resolve

differences of opinion, review overall pathology interpretations within a study, or resolve and document a complicated issue.

Sensitivity: In the context of this guidance, refers to a biomarker, an assay for detection of a biomarker, or a diagnostic test using a biomarker. **Biomarker sensitivity** is the probability that a biomarker will report a specific morphologic change when the change is present. A sensitive biomarker also demonstrates changes after a relatively low amount of target tissue damage. **Analytical sensitivity** of the detection method for a given biomarker refers to the smallest quantity of biomarker that can be measured. For clarity, the term *sensitivity* should be qualified as being related to the biomarker itself (i.e., biomarker sensitivity) or the methodology used to measure the biomarker (i.e., analytical sensitivity). **Diagnostic sensitivity** is the probability that a person having a disease will be correctly identified by a clinical test (the number of true positive results divided by the total number with the disease, which is the sum of the numbers of true positive plus false negative results).

Specificity: Refers to a biomarker, an assay for the detection of a biomarker, or a diagnostic test using a biomarker. **Biomarker specificity** is the probability that the biomarker does not change in the absence of change in the tissue of interest. A biomarker with high specificity is reliably unchanged or stable in the absence of a defined morphologic change. **Analytical specificity** for the biomarker refers to the ability of the assay to measure one specific substance in the presence of other substances expected to be present. **Diagnostic specificity** is the probability that a person not having a disease will be correctly identified by a clinical test (the number of true negative results divided by the total number of those without the disease, which is the sum of the numbers of true negative plus false positive results).

Truth Standard: Any medical procedure or laboratory method or combination of procedures and methods that the relevant scientific or clinical community relies upon for diagnosis or other specific categorization of the studied tissue, animal, or person and that is accepted by FDA for this purpose in biomarker qualification. Ideal truth standards will have negligible likelihood of either a false positive or a false negative result (Zhou, Obuchowski, and McClish, 2011).

REFERENCES

Bailly V, Zhang Z, Meier W, Cate R, Sanicola M, Bonventre JV, 2002, Shedding of kidney injury molecule-1, a putative adhesion molecule involved in renal regeneration. J Biol Chem, 277(42):39739-39748.

Biomarkers Definitions Working Group, 2001, Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework, Clin Pharmacol Ther, 69(3):89-95.

Burkhardt JE, Pandher K, Solter PF, Troth SP, Boyce RW, Zabka TS, Ennulat D, 2011, Recommendations for the evaluation of pathology data in nonclinical safety biomarker qualification studies. Toxicol Pathol, 39(7):1129-1137.

Crissman JW, Goodman DG, Hildebrandt PK, Maronpot RR, Prater DA, Riley JH, Seaman WJ, Thake DC, 2004, Best practices guideline: toxicologic histopathology. Toxicol Pathol, 32(1):126-131.

Digital Pathology Association (DPA), http://www.digitalpathologyassociation.org. (last accessed December 18, 2015).

FDA draft guidance for industry, Center for Drug Evaluation and Research, *Bioanalytical Method Validation* (Revision 1). Available on the Internet at http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm.

FDA guidance for industry, Center for Drug Evaluation and Research, *E16 Biomarkers Related to Drug or Biotechnology Product Development: Context, Structure, and Format of Qualification Submissions*. Available on the Internet at http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm.

FDA guidance for industry and FDA staff, Center for Drug Evaluation and Research, *Qualification Process for Drug Development Tools.* Available on the Internet at http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm.

Holland T and Holland C (2011a), Unbiased histological examinations in toxicological experiments (or, the informed leading the blinded examination). Toxicol Pathol, 39:711-714.

Holland T and Holland C (2011b), Analysis of unbiased histopathology data from rodent toxicity studies (or, are these groups different enough to ascribe it to treatment?). Toxicol Pathol, 39:568-575.

Keenan C, Elmore S, Francke-Carroll S, Kemp R, Kerlin R, Pedadda S, Pletcher J, Rinke M, Schmidt, SP, Taylor I, Wolf DC, 2009, Best Practices for use of historical control data of proliferative rodent lesions. Toxical Pathol, 37(5):679-693.

Kondylis FI, Moriarty RP, Bostwick D, Schellhammer PF, 2003, Prostate cancer grade assignment: the effect of chronological, interpretive and translation bias. J Urol, 170 (4 Pt 1):1189-1193.

Mignini LE, Khan KS, 2006, Methodological quality of systematic reviews of animal studies: a survey of reviews of basic research. BMC Med Res Methodol, 6:10.

Morton D, Sellers RS, Barale-Thomas E, Bolon B, George C, Hardisty JF, Irizarry A, McKay JS, Odin M, Teranishi M, 2010, Recommendations for pathology peer review. Toxicol Pathol, 38(7):1118-1127.

Piper RD, Cook DJ, Bone RC, Sibbald WJ, 1996, Introducing Critical Appraisal to studies of animal models investigating novel therapies in sepsis. Crit Care Med, 24(12):2059-2070.

Pound P, Ebrahim S, Sandercock P, Bracken MB, Roberts I, Reviewing Animal Trials Systematically (RATS) Group, 2004, Where is the evidence that animal research benefits humans? BMJ, 328(7):514-517.

Roberts I, Kwan I, Evans P, Haig S, 2002, Does animal experimentation inform human health care? Observations from a systematic review of international animal experiments on fluid resuscitation. BMJ, 324(7335):474-476.

Tuomari DL, Kemp RK, Sellers R, Yarrington JT, Geoly FJ, Fouillet XL, Dybdal N, Perry R, 2007, Society of Toxicologic Pathology position paper pathology image data: compliance with 21CFR Parts 58 and 11, Toxicol Pathol, 35(3):450-455.

Wei B, Jin JP, 2011, Troponin T isoforms and posttranscriptional modifications: Evolution, regulation and function. Arch Biochem Biophys, 505(2):144-154.

Zhou XH, Obuchowski NA, McClish DK (2011), *Statistical methods in Diagnostic Medicine*. Wiley and Sons, Inc, New Jersey.