Published in final edited form as:

*Toxicol Pathol*. 2017 January ; 45(1): 238–247. doi:10.1177/0192623316672352.

# Is it adverse, non-adverse, adaptive or artifact?

**Arun R. Pandiri**[1], **Roy L. Kerlin**[2], **Peter C. Mann**[3], **Nancy E. Everds**[4], **Alok K. Sharma**[5], **L. Peyton Myers**[6], and **Thomas J. Steinbach**[7]

[1]National Toxicology Program, Research Triangle Park, North Carolina [2]Drug Safety Research and Development, Pfizer Inc., Groton, Connecticut [3]Experimental Pathology Laboratories, Inc., Northwest, Seattle, Washington [4]Amgen Inc., South San Francisco, California [5]Covance Laboratories, Madison, Wisconsin [6]US Food and Drug Administration, Silver Spring, Maryland [7]Experimental Pathology Laboratories, Inc., Durham, North Carolina

## Abstract

One of the principal challenges facing a toxicologic pathologist is to determine and differentiate a true adverse effect from a non-adverse or an adaptive response. Recent publications from the Society of Toxicologic Pathology (STP) and the European STP provide guidance for determining and communicating adversity in nonclinical toxicology studies. In order to provide a forum to inform and engage in a discussion on this important topic, a continuing education (CE) course was held during the 2016 STP Annual meeting in San Diego, CA. The lectures at this course provided guidance on determining and communicating adversity using case studies involving both clinical pathology and anatomic pathology. In addition, one talk also focused on data quality, study design and interpretation of artifacts that could hinder the determination of adversity. The CE course ended with a talk on understanding adversity in preclinical studies and engaging the regulatory agencies in the decision making process. This manuscript is designed to provide brief summaries of all the talks in this well received CE course.

### Keywords

Adversity; Artifact; Adaptive response; Non-adverse response

## Introduction

One of the main purpose of nonclinical toxicology studies is to identify doses that correspond to “no observed effect level” (NOEL) and “no observed adverse effect level” (NOAEL). The guidance to identify NOELs are relatively well established and the parameters to consider include dose response, precision of the measurements, range of the natural variation (controls), the biological plausibility of the effect(s), statistical significance (ECETOC, 2002). The literature also provides guidance to determine NOAELs but this is subject to expert judgment, and guided by several criteria to be considered in identifying

Correspondence to: Arun Pandiri, National Toxicology Program, Cellular and Molecular Pathology branch, 111 TW Alexander Dr, Rall Building - Rm: B350, Research Triangle Park, NC 27709, pandiriak@niehs.nih.gov.

adverse, non-adverse and adaptive findings. In addition, care should be taken to minimize artifacts or systematic errors in the study design and data analysis that could complicate the decision making process of arriving at NOAELs.

Several recent publications aimed to provide guidance in determining adversity in preclinical toxicology studies (ECETOC, 2002, Lewis *et al.*, 2002, Dorato and Engelhardt, 2005, Keller *et al.*, 2012, Kerlin *et al.*, 2016, Palazzi *et al.*, 2016). The readers are encouraged to refer to these publications for more in depth information on adversity in toxicology studies. Due to the increased interest in the topic of adverse versus adaptive responses in toxicology studies, several recent workshops, and symposia were organized by the American College of Toxicology (Cavagnaro and Beilke, 2014, Kiorpes and Pandiri, 2014), the Society of Toxicologic Pathology (Kerlin *et al.*, 2016), and the European Society of Toxicologic Pathology (Palazzi *et al.*, 2016). This continuing education course at the 2016 STP annual meeting in San Diego, CA, aims to provide a forum to disseminate information and to enable discussion within the STP membership.

In the following sections, brief summaries of the lectures presented in the 2016 STP CE course “Is it adverse, adaptive or artifact?” are presented. In the first lecture, Dr. Roy Kerlin presented some key points for determining and communicating adversity in toxicological studies. Dr. Peter Mann presented an overview on artifacts in toxicologic pathology that need to be controlled in data generation before the data is considered for adversity considerations. Drs. Nancy Everds and Alok Sharma discussed clinical and anatomic pathology perspectives, respectively, in determining adversity in toxicology studies. Finally, Dr. Peyton Meyers provided a regulatory perspective on adverse findings in toxicology studies and how to effectively engage and communicate with the regulatory agencies.

## I. What is an Adverse Effect in Toxicologic Pathology?

Dr. Roy Kerlin (Pfizer, Inc.) presented an overview of the recommendations of the STP working group on adversity. The concept of the “No Adverse Effect Level” (NOAEL), defined by decisions to mark test article effects in a nonclinical toxicity study as “adverse or non-adverse”, has underpinned the fundamental framework of toxicity assessments and reporting, as well as risk management strategies for pharmaceutical agents for the last 4 decades. Because of this the STP, through the auspices of its Scientific and Regulatory Policy Committee, sponsored the formation of a Working Group to address this, led by the Author, and including Brad Bolon (GEMpath Inc.), John Burkhardt (Abbvie; Pfizer, from Jan 2016), Sabine Francke (U.S. Food and Drug Administration), Peter Greaves (Leicester Royal Infirmary, UK), Vince Meador (Covance), and James Popp (Stratoxon). The resulting publication cited below was endorsed by all the major societies representing the practice of Toxicologic Pathology supporting pharmaceutical development across the world: *Recommended (“Best”) Practices for Determining, Communicating, and Using Adverse Effect Data from Nonclinical Studies* (Toxicologic Pathology 2016 Feb;44(2):147-62).

This publication defined ten major recommendations to guide company authors and regulators about how best to establish adversity, communicate adversity, and use the concept consistently in developing strategies to develop new pharmaceutical agents. Each of these

will be cited as it is in the original publication with a brief discussion of what the recommendation means. In addition, a specific focus on adaptive effects under these recommendations will be incorporated into the discussion.

### 1. "Adversity" is a term indicating "harm" to the test animal, within the constraints of the study design (dose, duration etc.)

Although this term "harm" may seem simplistic, it extends to include many things that previous definitions which defined adversity on the basis of functional disruption excluded. For example, the degeneration of neurons in a grey matter nucleus with no known function may not be considered adverse under most of the previous attempts to define adversity in the literature, since there may be no measureable effect on animal function or ability to maintain homeostasis etc. However, it is clear that necrosis of an entire structure in the brain would be an adverse effect, and definitely be considered "harmful' to the animal under test. The important feature of this recommendation is that, where unclear, the nature of the "harm", or lack thereof in the case of a non-adverse effect, can be defined within the context of the study which produced the toxicity. Another important note regarding this definition is that reversibility *per se* is insufficient to establish lack of adversity, since many effects that do not damage infrastructure (such as some kinds of necrosis in a parenchymal organ) and are completely reversible, may be adverse. Similarly, the presence of biomarkers of the effect, while being an important aspect to highlight in a report, and which may assist developmental decisions for the agent under test, are not a predicator of the adversity decision.

### 2. The decision about whether or not test article-related effects (or a group of related effects) in a nonclinical study are considered "adverse" or "non-adverse" should be unambiguously stated and justified in sub-reports and/or the study report

While this recommendation also sounds simple, the application contains nuances and contextual variances that should be explained. What this effectively means is that any test article effect must be linked clearly to either an adverse or non-adverse decision. But what this does NOT mean is that every small test article effect, large or small has to be separated out and separately dealt with as adverse or non-adverse. Many adverse effects have a primary cause, such as myocardial necrosis, with secondary effects on liver and kidney due to cardiac insufficiency and perhaps clinical effects of inanition or syncope, and possibly changes in clinical chemistry parameters in peripheral blood such as cardiac troponin or aspartate aminotransferase (AST). As long as there is clear communication of the pathogenesis defined in the report, many of these effects can be lumped together as an adverse effect. For example, a paragraph from such a study may read, "Test article-related myocardial necrosis led to secondary congestion in liver and kidney due to cardiac insufficiency with increased levels of AST and troponin. These adverse effects were only observed at the high dose". Note that this does not eliminate the possibility that the liver or kidney effects may be due to a separate effect of the test article. But they are already lumped in as an adverse effect <u>at that dose</u> in any case, to make the most likely case for pathogenesis. This is a critical point that will be explored later: Presumptive, unproven pathogenesis should not be used to disregard a test article effect that would otherwise be considered adverse. Another example may be, "small changes (less than 0.95× control means) in Na, Cl, glucose and albumin observed only in males at the high dose were not

considered to be adverse due to their small magnitude, presence in only one sex despite similar exposure to the test article in females, and lack of any clinical or histological evidence of organ damage".

### 3. "Adversity" as identified in a nonclinical study report should be applied only to the test species and under conditions of the study

The implications of this recommendation are important. This simply says that interpretation of the adversity decision should be restricted to the test animal and the test system. The potential for extrapolating these results to the future may be critical for developmental decisions within a company but should not be used within a study report. For example, hyperplasia should be determined as adverse or not, based on what is currently observed (i.e. did the hyperplasia "harm the test animal?), not on the implications for this being a precancerous effect that might result in signals in a later carcinogenicity study. Nor should we determine adversity based on the potential to affect humans or not, because we did not test that within the animal study. Of course conjecture about human susceptibility to such effects is perfectly valid and even expected to be discussed within the report narrative, especially where the Pathologist is aware of human/animal correlates of a particular toxicity, or even where literature is available showing that certain target effects in animals have no counterpart in humans. However, this conjecture should not be used to predicate a decision for adversity for the nonclinical test animal within the study report. Of course with more data from multiple studies, it may build a "weight of evidence" case for or against the human relevance of a particular toxicity, but this case should be made within the nonclinical overview documents supporting human dosing where all of the information from individual animal studies is available.

Another point impinging upon human relevance is that adversity decisions should not be established differently depending on the indication for the drug. For example, we should not be more lenient with our adversity decisions for a study for a drug targeting cancer, than for a drug for a "lifestyle" indication. Indications change over time and a drug for cancer in older patients may be re-purposed for inflammatory conditions in pediatric patients. The reports should be able to be used within different overview documents supporting different indications. The alternative would be to re-issue reports with different adversity calls and different NOAELs for different indications, which would be logistically a nightmare, but also, it stretches the scientific credibility around what "adverse" really means, to the breaking point.

### 4. Toxic effects on cells, tissues, organs, or systems within the test animal should be assessed on their own merits

This recommendation is simply regarding speculation about the origin and/or pathogenesis of a test article effect. It says that authors should look at the effect alone and ask whether or not it caused harm to the test animal to assess adversity. There are three main aspects to this recommendation: i) Pathogenesis of the effect – primary or secondary; ii) Is the effect due to suprapharmacologic modulation of targets i.e. an exacerbation of efficacy; iii) Is the effect an exacerbation of an adaptive or background change.

**i. Pathogenesis of the effect—**It is sometimes tempting to establish a presumptive pathogenesis and assign adversity to the primary cause and ignore the speculative secondary or tertiary effects as irrelevant. While often authors can be accurate about such things it is a dangerous game to play. For example, lowered blood glucose levels may result in tremors and convulsions and may even create neuronal degeneration. However, it may also be true that the compound that creates hypoglycemia may, in addition, be a neurotoxin. In general, it is best to assign adversity on the basis of the actual effect, while still outlining pathogenesis in a report (and using that to make the report more easily readable as outlined earlier), but especially not using speculative pathogenic constructs to reject otherwise adverse effects. As long as the authors do not ignore an otherwise adverse effect on the basis of presumptive, but unproven pathogenic constructs, the lumping of multiple effects under the one adverse primary cause is a good way to simplify the report.

**ii. Suprapharmacology or exaggerated efficacy—**Sometimes authors of reports will want to ignore adversity decisions for toxicities determined to be a consequence of the expected pharmaceutical effects of the drug. The example outlined above could be cited again. A drug to treat diabetes may lower glucose to the extent that damage to neurons may occur with all the effects outline above. But to ignore this because we expect this and can manage it in the clinic is mistaken for two main reasons. Firstly, we can be wrong! As indicated above, the drug may also be a neurotoxin. Secondly it may actually be correct, but we do not know if animals are more or less sensitive than humans, and we may need to be especially careful in the clinic. Understanding that an adverse effect due to exaggerated pharmacology may occur in animal species (and the exposure margins at which it may occur) indicates that special attention should be paid to prevent these effects in human subjects and patients.

**iii. Exacerbation of adaptive or background effects—**It may also be tempting to call all adaptive effects or exacerbations of background effects that may have a species predilection non-adverse. And although there is a lot of discussion about this whenever the topic of adversity comes up, in reality most of such effects in species used for nonclinical toxicity studies are objectively non-adverse anyway (i.e. do not cause harm). But in the rare instances where such effects become adverse (such as when chronic progressive nephropathy (CPN) in rats develops to the extent it results in uremia, mineralization or even death) it should be called as such, even if it carries an explanation that it may not be relevant to humans. The main reason is that we usually do not understand the reasons why the exacerbation takes place. Authors may speculate that general debilitation at the highest dose caused the exacerbation, but they do not usually attempt to address why other studies in their facility with debilitated animals did not show a similar effect on that background change. Perhaps in the case of the rat CPN situation, the test article is actually a mild renal tubular toxin, but which may be expressed in rats by exacerbating this background change. Subsequent studies, perhaps in other species, may determine whether or not the material is really nephrotoxic. In a similar manner, an adaptive change may become adverse in the nonclinical species, and this may well be a signal that it could be a hazard for humans undergoing the same adaptation to the test article.

### 5. Communication of what is considered “adverse” and assignment of the NOAEL in the overall study report should be consistent with, and supported by, the information provided in the study sub-reports

The main aspect of this recommendation is the need to document all of the test article effects as either adverse or non-adverse within subreports (such as clinical pathology or anatomic pathology report) or the main report, but ensuring that the message is completely consistent between subreports and the main report. This does NOT mean that every finding in a subreport must have an adversity designation, since in many cases (such as clinical pathology effects) there may not be sufficient information available to the authors to make this designation at the time of writing. But it does mean that all the loose ends are tied up in the overall study report, such that somewhere in the totality of the reporting (sub report or main report), all test article effects are linked with either adverse or non-adverse decisions. And the other main point is that the NOAEL should be set ONLY in the overall study report, since it would not make sense to have different NOAELs for different subreports and the main report.

A last consideration for this recommendation is to comment on the use of terms such as “not biologically relevant” or “not toxicologically important”. While such terms can be acceptable, they are often ambiguous, and should be strictly defined. In general, their use should be limited, and should NOT be used as a surrogate for determining adversity/non-adversity, in order to ease the responsibility for making that designation in difficult cases.

### 6. Communication of adverse findings and the NOAEL should include direct interaction between staff within different contributing scientific disciplines

As stated previously, a single toxicity may manifest effects in a number of different end-points which may be captured by different authors of subreports. All of these will have to be brought together and integrated in the overall main study report. It is very important that there is early and clear communication between such authors so that the various manifestations of toxicity due to a single effect may not be thought to be due to multiple different toxicities. For example, clinical syncope may be due to a variety of effects, such as neurological damage. However, if the test article caused cardiac infarcts, it may manifest with syncope, electrocardiogram abnormalities, increases in AST and troponin, maybe changes in other renal or hepatic biomarkers, and finally histopathology evidence of necrosis, and maybe thromboembolism (if the cause might be due to endothelial damage or clotting). In these cases, communication and discussion between authors of the different sections will be essential to integrating and telling the correct story, versus listing an unrelated set of effects in a report.

### 7. The NOAEL for a test article should be communicated in an overview document based upon data from multiple studies

This is an important aspect of defining a scientifically valid risk management strategy, once the hazards have been identified. It is possible that the findings in one study may mitigate the risk identified in another study. For example, a test article-related exacerbation of CPN in the rat may be less relevant if further studies show no effect in the kidneys of dogs. However, if subsequent work, based on the potential risk in kidneys as a safety target identified in the

literature, shows that dogs don't express the receptor for the target, but rats do, this may increase the potential for a target-based risk for humans. Other work, perhaps in monkeys, shown to also express this receptor, may either reinforce or reduce the risk concern for this effect in humans. Many other examples may be given, but overall promote the concept that risk management and assessment are properly done in the overview documents comparing effects in multiple studies in order to understand the full spectrum of potential hazards identified in the individual studies.

### 8. In order to place them in appropriate context, the use of NOAELs in data tables should be referenced to explanatory text

The preamble to the FDA GLP regulations references the fact that raw data for pathology is comprised of BOTH the data tables AND the pathology narrative. The reason for this is that histopathology data can easily be misinterpreted by someone simply reviewing data tables of diagnoses. A diagnosis of necrosis/inflammation could mean that a test article effect was primarily necrosis with associated inflammation considered secondary to the chemotactic properties of the cellular debris. Or it could mean that there was fulminant inflammation with necrosis as a secondary event to the massive activation of inflammatory cells. Such different interpretations may not be evident simply by reviewing the diagnoses. This is the first reason why looking at NOAELs in tables of diagnoses without the textual references and perspective is very unwise. The second reason is that the NOAEL alone for a study or for a compound in an overview document does not contain the essential information about relative importance of the different adverse events. As an extreme example, the NOAEL in the most sensitive species for a drug may be set by liver toxicity, which is reversible and can be easily tracked with biomarkers of hepatocellular damage. But at a higher dose there may be an adverse finding of greater importance such as retinal toxicity, which may have no good biomarker and which is irreversible. Although this extreme example would likely be caught by someone, it shows how simply looking at the NOAELs in a table may leave a reviewer open to making mistakes about the relative importance of test article effects, without also reviewing the narratives of the studies.

### 9. Nonclinical scientists, including toxicologists, pathologists, and other contributing subject matter experts who interpret data from nonclinical studies, should be active participants in assessing and communicating human risk

The scientists who evaluated the data from a study are the best ones to communicate the potential hazards and thereby assist in evaluating potential human risk. This especially applies to the Study Pathologists, who are providing an interpretative opinion embodied in their diagnoses and report narrative. In the event that the original study Pathologists or other scientists are unavailable, such as in the case of reports generated at contract research organizations, or if the scientist leaves their organization, another suitably experienced scientist with similar credentials may be engaged to provide this service. In addition to providing appropriate interpretations for diagnostic terminology unfamiliar to other scientists, pathologists provide the advantage of bridging the clinical-nonclinical interface, since most pathologists are also veterinarians (or occasionally human physicians) and can assist human clinicians to understand the potential relevance and severity/impact to humans of various lesions in animals.

### 10. All available data from all nonclinical studies must be evaluated together to define any potential toxicities and to predict human risk

This recommendation is to reinforce that as data is gathered and further studies are run, the risk management strategy may change. This is obviously the case for nonclinical studies, since many hazards carry potential implications that may or may not be realized until tested. For example, there may be effects in seminiferous tubules that are shown subsequently to have no effect on male fertility in rats, etc. But this also comprises the results of clinical studies, which may indicate that an animal toxicity does not manifest in humans, or alternatively may appear at much lower drug exposure (i.e. humans are more sensitive to the effect). In addition, as time passes, the literature may reveal new insights, or experimental mechanistic studies may be run to prove or disprove various pathogeneses.

**Summary—**The Society of Toxicologic Pathology sponsored a Working Group to develop a series of recommendations about identifying, communicating and using adversity, or lack thereof, to define the importance of test article effects in nonclinical toxicity studies. The intent was to reduce inconsistences and maximize understanding between report authors and clinicians and regulators utilizing such studies to design safe and effective human clinical trials. The ten recommendations published recently will assist both authors and regulators to more confidently and consistently define adverse effects and, importantly, the NOAELs that help guide human dose selection. Presumptive adaptive test article effects should be considered adverse or non-adverse, as with any other test article effects, and the importance of these can be better established after the results of multiple studies are used to write the nonclinical overview documents, containing the results of numerous studies with multiple species over various spans of time.

## II. Adverse or adaptive? No, it is an artifact

Dr. Peter Mann (EPL, Inc.) presented an overview on artifacts that may be introduced when conducting a toxicology study and how to minimize them. When making a determination of adversity in a nonclinical study, it is important to ensure findings are first test article-related and not due to artifact. Artifacts can be introduced at any phase of a study but most often they occur due to inappropriate experimental design, study protocol design, necropsy, tissue processing, or finally at data capture and interpretation.

Selection of the suitable experimental animal model and appropriate dose setting are essential for a proper toxicologic study design. Poor protocol design can introduce artifacts into a study in a number of ways. Examples include failure to include a control group which could lead to over or under diagnosing lesions in the treated groups, not examining the low and mid dose groups, failure to follow SOPs, and assigning inexperienced personnel to data collection.

There are numerous possibilities to introduce artifacts at the time of necropsy. Excessive or poor tissue handling techniques can affect the quality of the tissue sample. Autolysis begins immediately following death so any unnecessary delay in the fixation of tissue can have profound effects on the tissue quality. The choice of fixative depends on the downstream applications. For most tissues, 10% neutral buffered formalin (10% NBF) is sufficient but

when used on the eye or testes, 10% NBF can introduce artifactual change. This may result in the misinterpretation of normal histology as degeneration so for these tissues modified Davidson's is a better choice of fixative for routine histopathology examinations. Likewise, if electron microscopy is planned, using fixatives that contain glutaraldehyde is essential for optimal results. Immersion fixation is sufficient for most tissues (assuming proper tissue thickness and tissue:fixative ratio). However, some tissues, such as lung, should be optimally inflated with fixative to ensure proper fixation and separation of alveolar walls. Others, such as urinary bladder and gastrointestinal tract should be either inflated with fixative or opened to expose the mucosal surface to fixative or marked autolysis may result. Even so, the very process of inflation can introduce artifact. Overinflating the lungs can artifactually separate perivascular tissue and be incorrectly diagnosed as edema and in the gastrointestinal tract over inflation can appear as villous atrophy to an inexperienced pathologist.

The processing of tissue samples also offers many possibilities for artifacts to be introduced. A classic example is introduction of artifactual vacuolation in the white matter of the brain by alcohol fixation. When tissues are put in the processer they will generally sit in alcohol, which is the first stage of processing. Normally the processer is started at the end of the workday so that tissues can be processed overnight and be ready for embedding the following morning. Weekends and holidays can affect how long the tissue sits in alcohol awaiting the start of processing. When this delay is excessive the result can be artifactual vacuolation that appears in the white matter of the brain.

Another classic example of misdiagnosis of artifact leading to misinterpretation of study results was the dark neuron effect. Tissue processing or poor handling techniques can induce an artifactual change to neurons, which results in shrinkage of the neuronal cell body and darkening of the cytoplasm. An inexperienced pathologist may misdiagnose this change as degeneration or necrosis, producing inaccurate or misguided conclusions regarding the effects of the test article (Barone and Moser, 2004).

Artifacts may also arise in reporting the pathology data and these can be minimized by using generally accepted nomenclature, consistently applying these diagnoses to the same change while controlling for diagnostic drift, applying the *apriori* determined diagnostic criteria and severity scores, and defining the thresholds. Ideally, the diagnostic terminology for the individual animals and summary tables should be the same. However, sometimes the data from the individual animal tables may be combined for the summary tables using a different terminology and this process should be clearly defined and justified in the report to avoid confusion.

Finally, artifacts can be introduced at the time of pathology data input and interpretation. An example of potentially introducing artifact at data entry is the use of a global entry key for not remarkable, sometimes referred to as a “hot” button. The inappropriate use of this button can give the false impression that entire groups of animals were examined when they were not.

In conclusion, artifacts may be introduced at various time points in a study. Many artifacts, if not detected, can be extremely confounding but with experience most can be recognized and correctly interpreted.

## III. Clinical Pathology Parameters and Establishing a NOAEL

Dr. Nancy Everds (Amgen, Inc.) presented on adverse versus adaptive effects in clinical pathology. In general, it is rare for a clinical pathology marker to be adverse in the absence of any other change. Typically, clinical pathology provides context for adverse findings associated with morphologic lesions or clinical signs. Examples of potentially adverse clinical pathology findings include low calcium values resulting in tremors, low peripheral blood cell counts that result from bone marrow necrosis, and alterations in serum sodium, chloride, and proteins as the result of gastrointestinal or renal toxicity. Wherever appropriate, examples from the Common Terminology Criteria for Adverse Events (CTCAE, 4.03) for humans were used as reference points to provide translational context (NCI, 2009).

The discussion on adversity in hematology values focused on red cell mass, leukocyte counts, and hemostasis. Red blood cell count, hemoglobin, and hematocrit, are the most important parameters to consider when making a determination of adversity with respect to changes in red cell mass. It is estimated that decreased oxygen delivery does not occur until there is an approximate 30% decrease in red cell mass. Interestingly, the CTCAE defines a grade 1 adverse event as a hemoglobin value that is less than 10 g/dL, which is about 30% lower than the average hemoglobin in human populations. Reticulocyte count is critical in determining if a change in red blood cell mass is adverse. A decrease in red cell mass accompanied by a robust and appropriate reticulocyte response is less concerning than the same change in red cell mass without a reticulocyte response. An appropriate reticulocyte response depends on a number of factors and varies with species, age, number of blood collections, and intercurrent conditions such as inflammation and inanition. Additional factors to consider when determining adversity of red cell mass changes include: magnitude of the effect relative to controls or predose levels, the rate of decline (20% decrease over a day is likely worse than 20% over 3 months), and temporal pattern of change-was the change transient or sustained with continued dosing.

Determination of adversity with changes in leukocytes depends on the type of leukocyte. For neutrophils, the adversity of decreased counts in cynomolgus monkeys and dogs is viewed similar to that in humans in that low neutrophil counts can cause increased risk of infection. Rats and mice, however, already have low neutrophil counts and spontaneous infections are not generally observed in these species even when neutrophil counts approach zero. The number of lymphocytes ranges broadly in normal health, and it is generally difficult to assess adversity based on cell counts. Typically, changes in lymphocyte numbers are the result of a specific target of the test article (e.g. T-cell inhibition). In many instances, tests other than hematology, such as T-dependent antigen response (TDAR), may be more appropriate than cell counts in determining adversity. Change in the other leukocytes, such as eosinophils and monocytes, generally do not drive adversity determinations.

The discussion on hemostasis focused on insufficient coagulation factors and low platelet counts. Clotting time assays (PT, APTT) are generally fairly sensitive to changes in coagulation factors under the conditions of a toxicology study, and prolongation of these tests is often not associated with any adverse outcome. Similarly, within the confines of the protected environment of most toxicology studies, cynomolgus monkeys and dogs can have profoundly low platelet counts with no spontaneous hemorrhage. Rodents normally have very high platelet counts (700-1500K/μL). In rodents, low platelet counts are variably associated with spontaneous hemorrhage. Because low platelet counts may not always manifest clinically or morphologically with spontaneous hemorrhage, these changes can be difficult to anchor in adverse events. Regardless, test article-related profound decreases in platelet counts are a concern for translational medicine.

Clinical chemistry findings were grouped into three general categories. Firstly, there are markers that are tightly regulated and can be adverse if homeostasis is not maintained. These included glucose, calcium, and electrolytes such as sodium, chloride, and potassium. Decreased glucose is often observed in toxicology studies as a pharmacologic effect of the compound. In general, monkeys and rodents can have lower glucose levels without showing clinical signs, whereas dogs are more similar to humans in their sensitivity to low glucose. When anticipating lowered glucose levels, the study design can incorporate food supplements to mitigate the effects. Changes in electrolytes are generally considered adverse when accompanied with significant downstream changes such as acidosis or alkalosis, neurological signs, muscle weakness, or ECG abnormalities. Secondly, there are markers that are not as tightly regulated but still must be maintained between a specific range to maintain health such as serum proteins, triglycerides, and cholesterol. Changes in these values, unless quite large, are generally not adverse within the confines of a toxicology study, but may have implications for human health. Finally, there are markers that do not cause adverse downstream effects, but are generally markers of adverse events. Examples of these types of parameters include most of the serum enzyme markers (ALT, AST, ALP, and CK), and bilirubin. These endpoints often can be correlated with underlying pathology, and are useful in that they are usually translational to human medicine.

In conclusion with respect to clinical pathology, each institution should develop consistent and defensible approaches to determining adversity. Clinical pathology parameters are rarely adverse by themselves and typically do not determine NOAELs but they support the determination of the NOAELs based on morphologic lesions or clinical signs. The determination of adversity of clinical pathology findings must consider all available data and should be pathogenesis-based, rather than focused on individual parameters.

## IV. Adaptive, Non-Adverse, and Adverse Responses in Non-Clinical Studies

Dr. Alok Sharma (Covance Laboratories) reiterated the definitions of adaptive, non-adverse and adverse responses to emphasize that adaptive responses are not always non-adverse and can become adverse when severe enough as described by Palazzi et al., 2016. The adaptive nature of a pathologic finding must not be used to regard it as non-adverse. The bulk of this lecture was focused on common and classic examples of each of these responses in non-clinical studies.

Squamous metaplasia of larynx and trachea, and hepatocellular and thyroid follicular hypertrophy were shown as examples of adaptive responses. Irritant xenobiotics can cause squamous metaplasia in the trachea and/or larynx. The ciliated and columnar epithelium in the trachea and larynx is replaced by squamous epithelium, which likely provides protection against the irritant effect of xenobiotics. Minimal squamous metaplasia of the larynx is frequently observed as a response in inhalation studies. Based on the lack of any published evidence that laryngeal metaplasia in rodents is a preneoplastic finding, Osimitz et al. (2007) recommended not to use it as a toxicologic endpoint for risk assessment in subchronic inhalation studies (Osimitz *et al.*, 2007). An ESTP expert workshop has proposed to use severity of laryngeal metaplasia to decide if it is a non-adverse or adverse finding; focal minimal to slight laryngeal squamous metaplasia could develop as spontaneous or xenobiotic-induced findings and should be judged as non-adverse, whereas diffuse moderate to severe findings affecting multiple levels of larynx has a potential to result in laryngeal dysfunction; such cases should be considered adverse (Kaufmann *et al.*, 2009). Presence of concomitant degenerative/necrotic and/or hyperplastic changes should also lead to assessing laryngeal metaplasia as adverse.

Several examples of hepatocellular hypertrophy, non-adverse and adverse were presented. Hepatocellular hypertrophy and increased liver weights are observed as an adaptive response with xenobiotics causing cytochrome P450 induction or stimulation of peroxisome proliferator-activated receptor alpha (PPARα). It is thought that at minimal to slight severity and without any accompanying necrotic changes and increased liver enzyme activity, functional impairment does not occur and hepatocellular hypertrophy is considered non-adverse. On the other hand, hepatocellular hypertrophy with concomitant hepatocellular necrosis and pronounced liver enzyme activity may be considered adverse (Hall *et al.*, 2012, Maronpot *et al.*, 2010).

A few examples of xenobiotic-induced exacerbation of spontaneous or background findings were noted. In beagle dogs, incidence/severity of spontaneous idiopathic polyarteritis may be increased due to certain xenobiotics such as benzodiazepines and vasodilators (Clemo *et al.*, 2003). This exacerbation is usually considered adverse. Chronic progressive nephropathy (CPN) is a spontaneous/background finding in rats. A variety of test articles, such as those causing α2u globulin nephropathy can exacerbate incidence/severity of CPN. Increased incidence, especially of moderate to marked severity, is considered adverse as there is potential for impairment of renal function. Another example of exacerbation of a background finding is the increase in the incidence/severity of alveolar macrophages in inhalation studies. Minimal to slight increase in alveolar macrophages is considered adaptive and non-adverse. An STP alveolar macrophage working group recommended using a *no threshold* approach while recording this finding, while separately recording any accompanying changes, such as inflammation and epithelial hyperplasia, rather than lumping the changes together, as the presence of these changes will lead to assessment of this adaptive finding being considered adverse (Nikula *et al.*, 2014). Retinal degeneration due to light exposure is a common spontaneous finding in albino rodents. However, the finding can be seen as a xenobiotic-induced exacerbation (Yamashita *et al.*, 2016). In general, light-induced retinal degeneration is manifested as localized lesion and is limited to superior-temporal regions, especially when not in the advanced stage of lesion development;

whereas xenobiotic induced retinal degeneration is more diffuse (De Vera Mudry *et al.*, 2013). Within the context of the study, the xenobiotic-induced retinal degeneration is considered adverse.

Certain xenobiotics may cause pathological changes due to excessive modulation of the pharmacological target, which is more than that required to achieve optimum efficacy. Glucokinase activator compounds indicated for the treatment of type 2 diabetes caused profound and persistent hypoglycemia at high doses in non-clinical studies conducted in euglycemic animals. The toxicity is manifested as a neuronal degeneration in the central nervous system and axonal degeneration in peripheral nerves (Pettersen *et al.*, 2014). To investigate if the adverse nervous system findings resulted from exaggerated pharmacology of the compound, Tirmenstein et al. (2015) dosed normally hyperglycemic Zucker diabetic fatty rats with a glucokinase activator compound (Tirmenstein *et al.*, 2015). Based on the absence of any neuronal and axonal degeneration in this model, it was concluded that the toxicity observed in euglycemic animals was secondary to the exaggerated pharmacology of glucokinase activators. Nervous system findings in euglycemic animals will be considered adverse; however, changes are not expected to be seen in diabetic patients as they are normally hyperglycemic, and treatment with glucokinase activators may not cause profound and persistent hypoglycemia required to produce the findings in the central and peripheral nervous system.

Lack of human relevance or translatability does not make a finding non-adverse, as adversity should be called within the context of the study. Ettlin et al. (2010) demonstrated an unusual duodenal finding caused by an anticancer vascular endothelial growth factor (VEGF)-receptor inhibitor in a 26-week rat study. At termination, rats had epithelial hyperplasia and infiltration of mucosal glands into muscularis externa, serosa, and gastric wall. Based on invasive characteristic, the finding was initially diagnosed as malignant and clinical trials were put on hold. However, the finding partially reversed at the end of a 13-week recovery phase and an expert group assessed it as non-malignant and termed it adenosis with reversible hyperplasia; clinical trials were resumed (Ettlin *et al.*, 2010). The finding was limited to rats and not noted in mice and dogs. Inomata et al. (2014) also saw similar duodenal findings limited to rats following administration of a VEGF receptor tyrosine kinase inhibitor lenvatinib (Inomata *et al.*, 2014). Based on a time course evaluation, it was concluded that primary finding was Brunner's gland degeneration/necrosis, which likely led to inflammation and displacement of ectatic crypts into submucosa and muscularis externa. Based on the exclusive presence of the finding in rats and not in other non-clinical species, the finding was considered adverse in rats; however, it was not considered relevant to humans.

Several examples of phospholipidosis including those in lung, bile duct and skeletal muscle were discussed. Generally, phospholipidosis occurs without any concurrent target organ toxicity or dysfunction and the general consensus is that phospholipidosis is an adaptive response; however some regulators consider phospholipidosis as an adverse event since certain xenobiotics cause concurrent target organ toxicity and dysfunction (Reasor *et al.*, 2006).

## V. Regulatory Perspective on Adverse Versus Adaptive Responses in Toxicologic Pathology

Dr. Peyton Meyers (US FDA) presented several case examples from the FDA submissions dealing with adversity and also provided an insight into the regulatory decision making process. In nonclinical risk assessment, a NOAEL (No Observable Adverse Effect Level) is a common tool that is utilized as a benchmark for pharmaceutical decision making. The NOAEL is simply a value (based on the determination of 'adverse' findings in nonclinical studies) that helps translate nonclinical findings to a clinical trial by means of a FIH (first in human) dose. By definition, a FIH dose should have no human data to support the calculation for clinical dosing. Therefore, the nonclinical NOAEL is used to help establish a FIH dose. There are several definitions of a NOAEL in the regulatory parlance, but for selecting a FIH dose, the most common definition of a NOAEL (per FDA Guidance) is "the highest dose level that does not produce a significant increase in adverse effects in comparison to the control group" (CDER, 2005).

Inherent in the process of translating nonclinical to clinical data in the NOAEL is the difference between adverse, non-adverse, and adaptive responses. However, the definitions themselves will vary depending on which source is explored. There are multiple definitions for "adversity". Some authors (Lewis *et al.*, 2002) define adversity as a "biochemical, morphological or physiological change that either singly or in combination adversely affects the performance of the whole organism or reduces the organism's ability to respond to an additional environmental challenge". Other groups (IPCS, 2004) suggest that adversity is "change … that results in an impairment of functional capacity, an impairment of the capacity to compensate for additional stress, or an increase in susceptibility to other influences". Non-adverse has equal variance in the terminology where some authors suggest that the noted findings must "affect the general well-being, growth, development or life span of an animal" (Lewis *et al.*, 2002). However, other authors (Holsapple and Wallace, 2008) suggest that not only are biological changes but also responses to a xenobiotic challenges after the insult is required to suggest that adversity has occurred.

Establishing adversity into the NOAEL calculation incorporates both objective and subjective expert analysis. This decision is, in essence, an expert opinion regarding the interpretation of the toxicological findings as well as the known pharmacology of the product and is subject to "iterative interpretation as the toxicology profile develops" (Dorato and Engelhardt, 2005). As with the multiple definitions of "adversity", the NOAEL itself has multiple definitions beyond the FDA Guidance. One definition suggests that NOAEL must be based on a "response that is both statistically indistinguishable from the control outcome yet significantly different from the response observed at the 'lowest observable adverse effect level' (LOAEL)" (Calabrese and Baldwin, 1994). Other authors suggest that the NOAEL must be the "highest dose where no effect considered deleterious to the well-being of the animal is observed" (Holsapple and Wallace, 2008). Since the determination of the NOAEL is based on both data as well as expert opinion, it is no wonder that there is occasional debate among experts regarding study NOAELs or adverse findings during drug development.

In deciding whether an effect is adverse or adaptive, it is less likely to be considered adverse if the effects don't cause alterations in the tissues or organs, the effects are transient, or the effect is limited. Furthermore, if the effect is isolated, not a precursor to an established effect, secondary to other adverse effects, or is a consequence of the experimental model then it would likely not be considered adverse (Lewis *et al.*, 2002). Lastly, if an effect is determined to be adaptive then it likely would not be adverse. However, experts may disagree on a given effect being adverse compared to adaptive.

There are many examples of adaptive responses in nonclinical studies. Some adaptive responses are a consequence of the test system. For many nonclinical studies, however, this adaptation is a response to environmental variations and stresses, whether physical or chemical, in order to maintain normal function and survival (Lewis *et al.*, 2002). Adaptive responses are usually unrelated to the toxicity of the test substance itself. Some general examples may be: 1) liver enzyme induction after pharmaceutical administration (with limited liver enlargement) as a result of metabolic activity or 2) mild respiratory changes (mucous cell hyperplasia, macrophage accumulation, epithelial changes) after respiratory challenge. It should be noted that even adaptive responses can transition to adverse under a myriad of circumstances. Therefore, it is important to ensure that the rationale is clearly communicated to the regulatory agency as well as including supportive data and rationale as to why a finding is considered either non-adverse or adaptive. It is important to consider follow up plans (such as mechanistic studies, future toxicology study planning, or clinical monitoring) when a finding may be interpreted as adverse instead of adaptive. Furthermore, there are regulatory pathways during the IND as well as prior to the IND (the pIND process) which can be used to engage regulatory scientists to discuss nonclinical findings. Regardless, it is crucial to submit a clear and supported written rationale when proposing that particular finding(s) in nonclinical studies are adaptive instead of adverse. Planning ahead and engaging regulatory partners early will help streamline the process and will prevent undue delays in the development programs.

In summary, the NOAEL is an important tool for initial safety evaluation for FIH studies. Inherent in this process is the determination of “adverse” and “adaptive” findings from the supporting nonclinical toxicology studies. The determination of a finding to be adverse or adaptive is based on expert opinion and thus some findings are occasionally subject to different interpretations. If there are disagreements or likely different opinions on a toxicological finding, then enhanced communication will help streamline that discussion.

## References

Barone S Jr, Moser VC. The effects of perinatal tebuconazole exposure on adult neurological, immunological, and reproductive function in rats. Toxicol Sci. 2004; 77:183. [PubMed: 14704375]

Calabrese EJ, Baldwin LA. Improved method for selection of the NOAEL. Regul Toxicol Pharmacol. 1994; 19:48–50. [PubMed: 8159814]

Cavagnaro, JA.; Beilke, LD. What the AEL is the NOAEL?. American College of Toxicology, 35th Annual Meeting; Orlando, FL. 2014.

CDER. Guidance for Industry: Estimating the Maximum Safe Starting Dose in Initial Clinical Trials for Therapeutics in Adult Healthy Volunteers, 40CFR Part 261. US Government Printing Office; Washington, DC: 2005.

Clemo FA, Evering WE, Snyder PW, Albassam MA. Differentiating spontaneous from drug-induced vascular injury in the dog. Toxicol Pathol. 2003; 31 suppl:25–31. [PubMed: 12597428]

De Vera Mudry MC, Kronenberg S, Komatsu S, Aguirre GD. Blinded by the light: retinal phototoxicity in the context of safety studies. Toxicol Pathol. 2013; 41:813–825. [PubMed: 23271306]

Dorato MA, Engelhardt JA. The no-observed-adverse-effect-level in drug safety evaluations: use, issues, and definition(s). Regul Toxicol Pharmacol. 2005; 42:265–274. [PubMed: 15979222]

ECETOC. Technical Report No. Vol. 85. European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC); Brussels, Belgium: 2002. Recognition of, and differentiation between, adverse and non-adverse effects in toxicology studies.

Ettlin RA, Kuroda J, Plassmann S, Hayashi M, Prentice DE. Successful drug development despite adverse preclinical findings part 2: examples. J Toxicol Pathol. 2010; 23:213–234. [PubMed: 22272032]

Hall AP, Elcombe CR, Foster JR, Harada T, Kaufmann W, Knippel A, Kuttler K, Malarkey DE, Maronpot RR, Nishikawa A, Nolte T, Schulte A, Strauss V, York MJ. Liver hypertrophy: a review of adaptive (adverse and non-adverse) changes--conclusions from the 3rd International ESTP Expert Workshop. Toxicol Pathol. 2012; 40:971–994. [PubMed: 22723046]

Holsapple MP, Wallace KB. Dose response considerations in risk assessment--an overview of recent ILSI activities. Toxicol Lett. 2008; 180:85–92. [PubMed: 18588961]

Inomata A, Nakano-Ito K, Fujikawa Y, Sonoda J, Hayakawa K, Ohta E, Taketa Y, Van Gessel Y, Akare S, Hutto D, Hosokawa S, Tsukidate K. Brunner's gland lesions in rats induced by a vascular endothelial growth factor receptor inhibitor. Toxicol Pathol. 2014; 42:1267–1274. [PubMed: 24499803]

World Health Organization, I.P.o.C.S. , editor. IPCS. International Programme on Chemical Safety (IPCS) Risk assessment terminology. United Nations Environment Programme, the International Labour Organization, and the World Health Organization, and produced within the framework of the Inter-Organization Programme for the Sound Management of Chemicals; Geneva: 2004.

Kaufmann W, Bader R, Ernst H, Harada T, Hardisty J, Kittel B, Kolling A, Pino M, Renne R, Rittinghausen S, Schulte A, Wohrmann T, Rosenbruch M. 1st international ESTP expert workshop: “Larynx squamous metaplasia”. A reconsideration of morphology and diagnostic approaches in rodent studies and its relevance for human risk assessment. Exp Toxicol Pathol. 2009; 61:591–603. [PubMed: 19285845]

Keller DA, Juberg DR, Catlin N, Farland WH, Hess FG, Wolf DC, Doerrer NG. Identification and characterization of adverse effects in 21st century toxicology. Toxicol Sci. 2012; 126:291–297. [PubMed: 22262567]

Kerlin R, Bolon B, Burkhardt J, Francke S, Greaves P, Meador V, Popp J. Scientific and Regulatory Policy Committee: Recommended (“Best”) Practices for Determining, Communicating, and Using Adverse Effect Data from Nonclinical Studies. Toxicol Pathol. 2016; 44:147–162. [PubMed: 26704930]

Kiorpes, AL.; Pandiri, AR. Differentiating Adverse from Adaptive Changes in Toxicology. American College of Toxicology, 35th Annual Meeting; Orlando, FL. 2014.

Lewis RW, Billington R, Debryune E, Gamer A, Lang B, Carpanini F. Recognition of adverse and nonadverse effects in toxicity studies. Toxicol Pathol. 2002; 30:66–74. [PubMed: 11890477]

Maronpot RR, Yoshizawa K, Nyska A, Harada T, Flake G, Mueller G, Singh B, Ward JM. Hepatic enzyme induction: histopathology. Toxicol Pathol. 2010; 38:776–795. [PubMed: 20585142]

NCI. Common Terminology Criteria for Adverse Events v4.03 (CTCAE). 2009. Retrieved Aug 15 2016, from National Cancer Institute (NCI), U.S Department of Human and Health Services, National Institutes of Health: http://evs.nci.nih.gov/ftp1/CTCAE/CTCAE_4.03_2010-06-14_QuickReference_5x7.pdf

Nikula KJ, McCartney JE, McGovern T, Miller GK, Odin M, Pino MV, Reed MD. STP position paper: interpreting the significance of increased alveolar macrophages in rodents following inhalation of pharmaceutical materials. Toxicol Pathol. 2014; 42:472–486. [PubMed: 24178583]

Osimitz TG, Droege W, Finch JM. Toxicologic significance of histologic change in the larynx of the rat following inhalation exposure: a critical review. Toxicol Appl Pharmacol. 2007; 225:229–237. [PubMed: 17991503]

Palazzi X, Burkhardt JE, Caplain H, Dellarco V, Fant P, Foster JR, Francke S, Germann P, Groters S, Harada T, Harleman J, Inui K, Kaufmann W, Lenz B, Nagai H, Pohlmeyer-Esch G, Schulte A, Skydsgaard M, Tomlinson L, Wood CE, Yoshida M. Characterizing "Adversity" of Pathology Findings in Nonclinical Toxicity Studies: Results from the 4th ESTP International Expert Workshop. Toxicol Pathol. 2016; 44:810–824. [PubMed: 27102650]

Pettersen JC, Litchfield J, Neef N, Schmidt SP, Shirai N, Walters KM, Enerson BE, Chatman LA, Pfefferkorn JA. The relationship of glucokinase activator-induced hypoglycemia with arteriopathy, neuronal necrosis, and peripheral neuropathy in nonclinical studies. Toxicol Pathol. 2014; 42:696–708. [PubMed: 24771080]

Reasor MJ, Hastings KL, Ulrich RG. Drug-induced phospholipidosis: issues and future directions. Expert Opin Drug Saf. 2006; 5:567–583. [PubMed: 16774494]

Tirmenstein M, Horvath J, Graziano M, Mangipudy R, Dorr T, Colman K, Zinker B, Kirby M, Cheng PT, Patrone L, Kozlosky J, Reilly TP, Wang V, Janovitz E. Utilization of the Zucker Diabetic Fatty (ZDF) Rat Model for Investigating Hypoglycemia-related Toxicities. Toxicol Pathol. 2015; 43:825–837. [PubMed: 26085543]

Yamashita H, Hoenerhoff MJ, Peddada SD, Sills RC, Pandiri AR. Chemical Exacerbation of Light-induced Retinal Degeneration in F344/N Rats in National Toxicology Program Rodent Bioassays. Toxicol Pathol. 2016; 44:892–903. [PubMed: 27230502]