

Review Article

Open Access

GLPs, Human Error, and Deviations: When Are Quality and Integrity Compromised?

David V. Gauvin^{1*}, Scott E. Boley², Noel D. Horton¹ and Theodore J. Baird¹

¹Drug Safety Evaluation, MPI Research-A Charles River Company, Mattawan, MI 49071, USA

²Toxicology Sinclair Research Center, 562 State Road DD Auxvasse, MO 65231, USA

Abstract

The Good Laboratory Practice guidelines (GLPs) of the US Food and Drug Administration are primarily a process of bookkeeping to ensure that agreed procedures have been followed and procedural documentation is true and accurate. It is the goal of the GLPs to ensure that any study can be re-created through the accurate and detailed description contained within the audited final study report. The GLPs require the designation of a Study Director (SD) for each nonclinical safety study. Under the administrative policy, a SD represents "the single fixed point of responsibility for overall conduct of each study" (21 CFR 52 (172) 33770). The SD is charged with the technical conduct of the study including interpretation, analysis, documentation, and reporting of the results. As such, when an error or deviation is made on a study it is the SD, alone, that must ensure accurate and detailed description of the error, and initiation appropriate 'due diligence' to ensure that similar events on the study are minimized, and that the final report contains a clear and concise listing of all errors and guideline deviations, as well as a fair assessment of their potential impact on the overall quality and integrity of the data. Deviations happen on studies conducted in the best of laboratories, this review details a process of remediation that must take place to ensure that the study integrity is intact and need for repetition of the study is minimized.

Keywords: Good laboratory practice; Industry best practices; GLP compliance; Study errors

Abbreviations: GLPs: Good Laboratory Practice Guidelines; IBP: Industry Best Practices.

Introduction

Data generated during early nonclinical general toxicology and safety pharmacology studies are used to set a safe starting dose for first-in-human (FIH) clinical trials. A large majority of these data are generated in "for-hire" studies under contractual agreements with private or public research organizations (CROs). In early human clinical trials, the research subjects are typically healthy volunteers who have little to gain but much to lose if the drug is associated with unexpected, especially adverse, outcomes such as the TGN1412 (TeGenero[®]) or BIA 10-2474 (Bial[®]) incidents [1-3]. With that in mind, Good Laboratory Practice regulations (GLPs; 21 CFR, Ch. 13 §58.1) require a set of standardized nonclinical study procedures that are used to establish and document a process of data retention and produce a formalized report, including (anticipated and unanticipated) study-related events so that any single study can be thoroughly understood in the context in which it was conceived and executed, and reconstructed as needed. There is a subliminal, often unvoiced concern about the possibility that financial profits for contractual study data may contribute toward possible coercion or undue influence capable of distorting the judgment of members of the CROs who must balance their fiduciary responsibility to conduct valid and reliable research for submission to the FDA and the potential loss of a paying client as a result of program errors. The concept of undue influence from Sponsor-funded research is not trivial. Potential loss of a major client can lead to internal pressures to minimize the impact of study errors and the threat of regulatory oversight or audits following an adverse finding in subsequent clinical trials. In today's economy there is often a question of whether or not there is something uniquely distorting about money as opposed to a chance to participate in the development of a true medical cure. The financial rewards or loss of funding can create an impression of a conflict between two competing interests in the data-generation process of drug development. The amount of money needed for an Investigational New Drug "package" of studies can create a sense of economic influence on the decision-making process; the threat of the

disgruntled client leaving to pursue other laboratories provides even more pressure on the SD/researcher who is attempting to maintain neutral and assure study integrity. The GLPs do not guarantee study validity or reduce the likelihood of study errors being made. The GLP regulations merely provide an administrative policy to ensure a means to an end for the determination of a *post hoc* analysis of the true impact of deviations. The regulations require test facility management to assign a single point of control for each study – the Study Director (SD). Under the GLPs, the SD is not required to be a scientist, and the guidelines do not delineate any unique features required for SD status, except that they are selected by Test Facility Management based on their assessment of the SD's background, education, and experience [4]. Deviations from SOPs, protocol, or GLPs occur on study. The vast majority are not intentional or malicious, and they generally don't elicit a punitive response. There is no perfect study; it would seem fruitless to deny this fact knowing that there are so many "hands in the mix". Human deviation has been defined by Rodriguez-Perez [5] as:

a departure from acceptable or desirable practices on the part of an individual resulting in unacceptable or undesirable result.

According to Rodriguez-Perez [5] these deviations and mistakes are the symptoms of causal (human) factors associated with root causes that we must discover prior to solving them.

Documentation errors or simple clerical errors occur in every organization and bureaucracy. Contributing to these is the well characterized speed and accuracy trade-off characteristic of human performance [6]. With increasing workload and limited resources,

***Corresponding author:** David V. Gauvin, Department of Neurobehavioral Sciences, Drug Safety Evaluation, MPI Research, 54943 North Main Street, Mattawan, MI 49071, USA, Tel: +1 269 668 3336, E-mail: david.gauvin@mpiresearch.com

Received July 25, 2018; **Accepted** August 13, 2018; **Published** August 20, 2018

Citation: Gauvin DV, Boley SE, Horton ND, Baird TJ (2018) GLPs, Human Error, and Deviations: When Are Quality and Integrity Compromised? Pharmaceut Reg Affairs 7: 205. doi: [10.4172/2167-7689.1000205](https://doi.org/10.4172/2167-7689.1000205)

Copyright: © 2018 Gauvin DV, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

short-cuts and errors are going to happen. It is the responsibility of the Test Facility management to ensure adequate resources are available for each study undertaken at that particular site. Dr. James Reason [7] describes two approaches to the problem of human errors: the ‘person’ and the ‘system’ approaches. According to Reason [7] the “person approach” focuses on the errors of individuals, blaming them for forgetfulness, inattention, etc.; while the “system approach” concentrates on the environmental conditions under which individuals work and attempts to build “safety nets” to avert errors or mitigate their effects. Other organizations emphasize the philosophy that improvement is about learning. The reason deviations occur is often because something is misunderstood in the processes established by the organization. The pharmaceutical industry and drug regulators have attempted to “force-fit” corrective action plans (Lean Six Sigma, CAPA, etc.) that are useful to mitigate production errors in manufacturing processes to errors in human services or human processes of bench-top or animal laboratory research. One of the significant criticisms of such approaches is related to the insight that the mechanical steps of equipment assembly line production (e.g., as in GMP) may not be directly relevant to the dynamic properties of human behavior (e.g., as in GLP), particularly in complex and relatively more sophisticated non-manufacturing contexts. The conduct of human researchers performing preclinical safety assessments with whole animals (technicians, scientists, support staff, etc.) involves cognitive processes of attention, knowledge-based decision making, rule-based decision making, contextual orientation, and memory. Manufacturing production line errors either do not involve these processes, or do not engage requisite cognitive faculties to the same extent. Contract Research Organizations (CROs), even highly reliable ones, certainly experience and contribute to their share of mistakes and recognize that human variability in performance is a force to harness in averting errors. Under the GLP requirements “all hands” within the organization must be adequately trained for the jobs they perform. There are daily laboratory tasks that are so basic as to not require training, and others that require initial training with periodic and continual re-training. In the process of error analysis it may be discovered that ineffective training is a “cause” of errors in the study conduct. Human resources, time allocation, and equipment failures may also represent potential sources of error. Technicians may have insufficient practice-time or actual “hands-on” experience with a procedure that is infrequently employed. For example, specialized equipment or service offerings may be limited to a single therapeutic target that engages a dedicated study only every 2 to 3 years. How does a CRO maintain proficiency on these special assays in such circumstances? Despite the foregoing, it is worth remembering that just because an error has been made does not necessarily mean the critical study data on which pivotal decisions depend have been compromised. In spite of inevitable human errors, scientific integrity is not necessarily compromised. The SD is the centralized authority for study conduct. The SD is integral to study performance and is the single point of contact for all communications and events related to the performance and execution of the approved study protocol. In general, the SD is coordinating a relatively standardized set of protocols which are conducted repeatedly within the same institution for almost every new drug that is screened. For FDA to reject a study, it is necessary to find that there were deviations from the GLPs and that these deviations were of such a nature as to compromise the quality and integrity of the study covered by the agency inspection [8]. The term “reliability” refers to the inherent quality of a data parameter or set of parameters in a regulatory report submission relating to: 1) A clearly described experimental design to allow for the study to be reconstructed and repeated independently, if needed; 2) The methods intended for use on a study and those experimental procedures that were actually

performed; and 3) The reporting of the results to provide evidence of the reproducibility and accuracy of the findings. A well-documented study from initiation to termination can serve as a testament to the scientific integrity of the data and their interpretation. The opinion of the SD as to the validity, reliability and integrity of the completed study data hinges on a decision based on the totality of evidence and personal experience. If the GLP process remains intact during the conduct of the study, the acceptance of the study data by the SD is not necessarily a problematic task and should be reasonably easy to defend under formal, post-hoc auditing by the regulatory agency and Sponsors. The purpose of this review is to illustrate a general process by which the full impact of human and experimental errors can be viewed and summarized, in order to establish grounds for acceptance or rejection of the study data without undue influence of the management or the paying Sponsor. Regulators have the statutory responsibility to make sound and verifiable decisions and judge the reliability of each study that is submitted for review in accordance with scientific principles regardless of whether they were conducted in accordance with the GLP’s and/or standardized methods [9]. Based on the totality of available evidence from the study data and documentation, the SD must ensure that the quality and integrity of a study will hold up to regulatory scrutiny when the regulatory documents for a drug are submitted for review and approval.

Due diligence

When an error occurs on a study it is imperative that the SD is informed with limited delay. The SD must be given a full and detailed analysis of what was supposed to occur on the specific phase of the study under examination, what did occur, and a sound reason why there is a difference. As the single point of control, the SD must utilize descriptive information provided by the technical/operations staff to determine the relative magnitude and importance of the error, identify the source of the error, and initiate a plan to reduce the future likelihood of similar errors. The SD must also demonstrate sufficient understanding of the scope, magnitude, and practical implications of the deviation on the quality and/or integrity of achieving the objectives of the study protocol. Contemporaneously with the deviation on study, the SD must determine if this unplanned event in the study conduct represents a potential threat to the validity, reliability and integrity of the study data that is intended for federal regulatory review in support of the Sponsor’s future request for conducting clinical trials or obtaining a marketing permit of this regulated product. At the time of data submission to the FDA the regulatory staff must evaluate the effects of all errors, deviations or incidences of GLP non-compliance and:

1) Determine that the error or non-compliance does not affect the validity of the study and accept it, or

2) Determine that the noncompliance may have affected the validity of the study and require that the study be validated by the Sponsor submitting it, or

3) Reject the study completely.

The key component in this regulatory submission is the documentation of SD’s due diligence in addressing the issue at the time of the noncompliance occurred (43 FR 59989-59990). In the worst-case scenario, even the most egregious unintentional error may be useful to the Sponsor. In the preamble of the GLP regulations, the FDA (41 FR 51215) acknowledged that valid data and information in an otherwise unacceptable study which demonstrates an adverse effect of the product, may serve as the basis of the final regulatory submission. By

the FDA's own admission a technically bad study can never establish the absence of safety risk but may establish the presence of a previously unsuspected hazard (43 FR 59992). As described by Moermond et al. [9] the GLPs require that the protocol is fully documented, as are any deviations from the protocol, SOPs or GLPs and that all raw data are available. Coincident with the development of the GLPs was the development of standardized test guidelines, for instance, by the USEPA and OECD. These standard guidelines do not guarantee that the correct hypothesis, experimental design, or most appropriate species is tested. In addition, the established core battery of tests do not ensure that all relevant adverse responses for a given substance are tested [10] and they may be modified in protocol development to cover key issues specifically relevant to the test article and its therapeutic target [11-13]. However, results from non-standardized studies reported in peer-reviewed journals may, in some cases, contribute additional and important information to a risk assessment and should not necessarily be excluded from risk assessment simply because the study was not performed according to GLP and/or standardized guidelines. A peer-review study that followed nonstandard methods can be scientifically valid without GLP compliance; however, peer-review of these studies does not guarantee that the results are of sufficient quality [14,15]. The actions and documentation of the SD's due diligence in response to study errors or noncompliance are critical in the regulatory agency's decision-making process for marketing approval that may occur months to years from the actual calendar date of the individual study event in question. Administrative precedence has been established through collaborative and published risk assessment strategies from both industry and government regulatory agencies in both the US and Europe [11-18].

Establishing a data defense

For each deviation or incident of GLP noncompliance the SD must make a judgment as to the full impact of the event on the quality and integrity of the data from that day of the study. At the study report phase the SD must also review all deviations conducted on the study and make a global statement as to the overall impact of all incidences of study errors, deviations, and noncompliance. This can be accomplished by using a "data defense". The first step in determining what information is needed to substantiate a data defense is to understand the meaning of the claim of data integrity and/or reliability and to clearly identify each implied and documented deviation on the study. Risk analysis should identify "critical tasks" required by the protocol. Critical tasks are study directed functions that, if performed incorrectly or not performed at all, would or could compromise the study outcome. An example of a critical task would be an incident of unintended paravenous leakage from an intravenous dose administration. Extravasation injury is defined as the damage caused by the efflux of solutions from a vessel into surrounding tissue spaces during intravenous infusion. An infusate leaking into the subcutaneous tissues may be painful to the animal but some experimentally-induced metabolic states, such as diabetes, increases the likelihood of such incidents. The damage can extend to involve nerves, tendons, and joints and can continue for months after the initial insult. These findings are not directly test-article related, but the secondary effects to the metabolic disorder being examined. Risk analysis should also identify the "intended target" and/or the "expected environment". The dose group would be considered a significant intended target and the maintenance of room temperature, humidity, and light cycle would be considered the expected environmental conditions listed in the protocol that may have the potential of affecting the data collected on the study. According to documentation at FDA, the standard of competent and reliable scientific evidence has been

legally and administratively defined in federal case law of the Federal Trade Commission (FTC) as: [19] tests, analyses, research, studies, or other evidence based on the expertise of professionals in the relevant area, that has been conducted and evaluated in an objective manner by persons qualified to do so, using procedures generally accepted in the profession to yield accurate and reliable result. There is clearly no statutory or administratively established formula as to how many or what type of supportive evidence is needed to substantiate a claim of study data integrity. However, past history with regulatory reviewers has demonstrated that the agency will consider the "accepted norms" of the relevant research fields or disciplines of science and the documented consultations with experts from the various disciplines in rendering judgment. If there is an existing standard for substantiation developed by another government agency (i.e., CDC, EPA) or other authoritative body (i.e., Society of Toxicology (SOT), American College of Toxicology (ACT), etc.) the regulators generally will accord deference to those standards, as well. In determining whether the substantiation standard for data integrity have been met with competent and reliable scientific evidence, the SD should consider four issues in the assessment: The meaning of the claim that the quality and integrity of the data have not been compromised by the error or deviation:

1. The relationship of the evidence to the claim.
2. The quality of the evidence.
3. The totality of the evidence.

The first step in determining the "who, what, when, and where" of information that is needed to establish a data defense is whether the SD has an understanding of the relevance of the deviation to specific requirements of the protocol, a relevant SOP, or to specific sections of the GLPs. The SD must be able to identify each suspected and documented error or noncompliance event on the study. Initially, the SD should not limit the focus on individual deviations, but on what expected effect is being promoted when all of the statements being made over the full study are considered together. While it is important that the SD assesses each individual event (deviation or error), it is paramount to substantiate the overall "message" contained when the claims of data reliability are considered together. Included in the determination of data integrity, the SD must also consider whether the evidence has any relationship to the specific claim of compromised data being made or to the study interpretation itself. In determining if a single deviation has affected the quality and integrity of the data generated during that single event, the SD must consider if the overall scientific quality of the overall study has been compromised.

Scientific quality of a study is based on several criteria including:

The sample size (power analysis)

The study design, itself (are vehicle control cohorts present? Negative controls? Positive control?)

Data collection methods (validated computer software system? Validated and calibrated instrumentation?)

Statistical analysis? (will loss of animal subjects jeopardize power?)

The level of measurement for the dependent measures collected during the related deviation? – nominal, ratio, ordinal data? If the overall study adequately addresses all or most of these criteria, the study may be considered to retain its' high quality standard, in spite of the errors or incidents of noncompliance. The regulatory determination of acceptability will be made contingent upon the numerous scientific and statistical principles used as evidence to substantiate the data

defense. The Agency for Healthcare Research and Quality and Research (AHRQ) has defined “methodologic quality” as the extent to which all aspects of a study’s design and conduct can be shown to protect against systematic bias, nonsystematic bias, and inferential error [20]. The regulatory agencies hold quality to be the extent to which a study’s design, conduct, and analysis have minimized selection, measurement, and any number of additional potentially confounding biases. These criteria should be the first to be assessed when trying to integrate the post hoc analysis into the data defense. What deviations conducted during a study can influence selection or measurement bias? As a general principle the SD should think about the type of evidence that would be sufficient to substantiate a claim in terms of what “experts in the field” would consider to be valid and reliable. Competent and reliable scientific evidence to support a claim should first rely on data derived primarily from similar toxicology/safety studies conducted in the laboratory. The strongest nonclinical evidence is based on data from studies in identical animal models, on data that have been reproduced in the research laboratory and other research laboratories and on data that gives statistically significant dose-response relationships. With respect to within- and between laboratory data comparisons, the AHRQ recommends consideration of the following domains:

1. Comparability of subjects.
2. Exposure or intervention.
3. Outcome measurement (validated computerized systems? Calibrated instruments, observational study parameters?).
4. Statistical analysis (i.e., power analysis as it relates to loss of sample size).

As before, these apply to systems that represent acceptable approaches for assessing the quality of observational studies. For example, in determining the full impact of a dosing error on a study, one might calculate the change in “total body burden” over the course of the study to make a claim that a single or multiple events on a study had a minimal impact on the actual delivered dose for the individual animal or the study design dosing group, as a whole (for example, some SOP-derived threshold, such as less than 10% for formulated solutions). The SD must maintain an industry “best practice” approach in determining the data defense procedure being sure to take into account the types of study under review. The SD is admonished that systems used to rate the quality of both observational (subjectively derived data collection) studies and objectively derived data collection studies (i.e. ECG, quantitative clinical pathology parameters)—what we refer to as “one size fits all” quality assessments—may prove to be difficult to use and, in the end, may measure study quality less precisely than desired. The data defense evaluation should also consider 3 other domains—quality, quantity, and consistency. These are considered well-established variables for characterizing how confidently we can conclude that a body of knowledge provides information on which regulatory policy makers can act. As described by the AHRQ:

Quality is defined as the aggregate of quality ratings for individual studies, predicated on the extent to which bias was minimized; Quantity refers to the magnitude of effect, numbers of comparative studies, and sample size, or power; and The Consistency for any given analysis refers to the extent to which similar findings are reported using similar and different study designs. The SD should apply significant weight in his/her analysis to determine “competent and reliable scientific evidence” to:

Historical control data from the laboratory

Contemporaneous vehicle control data (from study cohorts), and

Published data from similar age- and strain-matched cohorts derived from standard reference books, published studies from peer-reviewed scientific literature, and

The background, education, and experience of the SD himself (herself).

Totality of evidence

Regulatory decisions regarding the validity, reliability, and integrity of study data are based on the “totality of evidence”. The FDA has adopted the “weight of evidence” approach as the standard in the regulatory review process for regulatory submissions. The “additional clarity,” as the agency calls it, expands the value of direct and circumstantial evidence as part of the “totality of evidence” that the decision-makers objectively intend to use to accept or reject the full impact of study deviations and noncompliance events on the reliability and integrity of the data. As the body of evidence grows, additional studies (i.e., quantity) conducted in the same laboratory increase the likelihood of a large range of quality scores and heterogeneity with respect to outcomes measured, and results that can be used to make the final judgment on data defense. When dose- and time-dependent related factors are similar across studies, consistency (and thus, strength of evidence) is enhanced. Differences in derived data reflect a reduction in consistency, and the regulatory reviewers will assume a diminution in the overall strength of the evidence. In the final analysis SDs attempting to grade the strength of the evidence supporting the integrity of the data and quality and reliability of the study report should always rely on value of an independent scientific review of the final data set as part of standard institutional practices. In judging the strength of the body of evidence in a single study used to make a final decision on the status of the overall study, the agreement of a “legally competent authority” is invaluable to regulatory decision-makers. A legally competent authority (LCA) is judged on the specific task at hand. The determination of who is an expert in a discipline of science is generally based on:

The established reputation of the scientist within the relative discipline (membership in scientific organizations, societies, and/or guilds),

A documented publication history that demonstrates the capacities to reason and deliberate, hold appropriate values and goals of that scientific discipline,

A knowledge of the related regulatory agency’s administrative policies, and

Can appreciate the critical value of the circumstances, and understands the relevant information one is given and can communicate a learned opinion.

The broader agenda to be met by the laboratory review process is to apply a set of rating and grading schemes in ways that can be made transparent for the Sponsor and the FDA decision makers who use the audited study report for approval determinations for marketing licensure. By a full review and documentation of the decision making process used to judge the full impact of study deviations and noncompliance events related to the GLPs, it is the totality of evidence that should be conveyed in the report. As the single point of control on the study, the FDA must accept the SDs conclusion with “a reasonable scientific certainty” or to a “reasonable degree of certainty” of the

finalized study report data. The message should be conveyed with confidence. The intent of this review is to move the evidence-based practice field for this process ahead in ways that will bring benefit to the entire nonclinical safety assessment industry.

Acknowledgement

This manuscript represents a "work for hire" publication as part of the responsibilities of the authors as Study Directors in a Contract Research Organization. There are no financial or contractual conflicts of interest. All authors were actively employed by MPI Research (A Charles River Company) during the development and production of this manuscript.

References

1. Tyrsin D, Chuvpilo S, Matskevich A, Nemenov D, Römer PS, et al. (2016) From TGN1412 to TAB08: the return of CD28 super agonist therapy to clinical development for the treatment of rheumatoid arthritis. *Clin Exp Rheumatol* 34: 45-48.
2. Hünig T (2012) The storm has cleared: lessons from the CD28 super agonist TGN1412 trial. *Nat Rev Immunol* 12: 317-318.
3. Kaur R, Sidhu P, Singh S (2016) What failed BIA 10-2474 Phase I clinical trial? Global speculations and recommendations for future Phase I trials. *J Pharmacol Pharmacother* 7: 120-126.
4. Brock WJ (2014) The role of the study director in nonclinical studies: Pharmaceuticals, chemicals, medical devices and pesticides. Hoboken, NJ: John Wiley & Sons.
5. Rodriguez-Perez J (2011) CAPA for FDA-Regulated Industry. Milwaukee, WI: ASQ Quality Press.
6. van Veen V, Krug MK, Carter CS (2008) The neural and computational basis of controlled speed-accuracy tradeoff during task performance. *J Cogn Neurosci* 20: 1952-1965.
7. Reason J (2000) Human error: models and management. *BMJ* 320: 768.
8. <http://wayback.archiveit.org/7993/20170111215809/http%3A/www.fda.gov/ICECI/Inspections/NonclinicalLaboratoriesInspectedunderGoodLaboratoryPractices/ucm072738.htm>
9. Moermond C, Beasley A, Breton R, Junghans M, Laskowski R, et al. (2017) Assessing the reliability of ecotoxicological studies: An overview of current needs and approaches. *Integr Environ Assess Manage* 13: 640-651.
10. Beronius A, Molander L, Ruden C, Hanberg A (2014) Facilitating the use of non-standard *in vivo* studies in health risk assessment of chemicals: A proposal to improve evaluation criteria and reporting. *J Appl Toxicol* 34: 607-617.
11. Ågerstrand M, Breitholtz M, Ruden C (2011) Comparison of four different methods for reliability evaluation of ecotoxicity data: A case study of nonstandard test data used in environmental risk assessments of pharmaceutical substances. *Environ Sci Eur* 23: 17.
12. Ågerstrand M, Edvardsson L, Ruden C (2014) Bad reporting or bad science? Systematic data evaluation as a means to improve the use of peer reviewed studies in risk assessments of chemicals. *Human Ecol Risk Assess* 20: 1427-1445.
13. Ågerstrand M, Kuster A, Bachmann J, Breitholtz M, Ebert I, et al. (2011) Reporting and evaluation criteria as means towards a transparent use of ecotoxicity data for environmental risk assessment of pharmaceuticals. *Environ Pollut* 159: 2487-2492.
14. McCarty LS, Borger CJ, Mihaich EM (2012) Information quality in regulatory decision-making: Peer review versus good laboratory practice. *Environ Health Perspect* 120: 927-934.
15. Moermond C, Kase R, Korkaric M, Ågerstrand M (2016). CRED: Criteria for reporting and evaluating ecotoxicity data. *Environ Toxicol Chem* 35: 1297-1309.
16. <https://pdfs.semanticscholar.org/f25b/9578aa8bd83a68b2baf067606fad0efe6ace.pdf>
17. Klimisch H-J, Andreae M, Tillmann U (1997) A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. *Reg Toxicol Pharmacol* 25: 1-5.
18. vom Saal FS, Myers JP (2010) Good laboratory practices are not synonymous with good scientific practices, accurate reporting, or valid data. *Environ Health Perspect* 118: A60.
19. <https://www.fda.gov/food/guidanceregulation/guidancedocumentsregulatoryinformation/ucm073200.htm>
20. <http://c.ymcdn.com/sites/www.energypsych.org/resource/resmgr/imported/Systems%20To%20Rate%20The%20Strength%20of%20Scientific%20Evidence%20-%20AHRQ.pdf>