GOOD LABORATORY PRACTICE GUIDELINES – QUALITY ASSURANCE / QUALITY CONTROL CRITERIA: THE QUEST FOR ACCURACY AND RELIABILITY IN TRACE ANALYTICAL METHODOLOGY

T. J. Nestrick and L. L. Lamparski

The Dow Chemical Company
Michigan Division Research and Development
Analytical Sciences, Special Analysis, 1602 Building
Midland, Michigan 48674 USA

ABSTRACT

Increasing concerns regarding product quality and environmental protection have recently impacted analytical laboratories via governmentally mandated Good Laboratory Practice (GLP) guidelines and associated Quality Assurance-Quality Control (QA/QC) procedures. While these operating criteria are designed and implemented to ensure reasonable accuracy and reliability for the measurement of specific analytes, they do not necessarily always achieve their goal. Often GLP and QA/QC criteria relate to specific methodology and instrumentation rather than to analyst capability, procedural weaknesses and problems related to the application of equivalent techniques. Therefore, analytical chemists must now concern themselves not only with mandated GLP and QA/QC, but also with those aspects of their methods which truly govern data accuracy and reliability. Detailed examination of every step within a given method, and the understanding derived therefrom, is the simplest route to achieving analyses that are simultaneously good, fast and cheap.

INTRODUCTION

Governmental agencies charged with the responsibility of assessing analytical data associated with manufactured product quality and environmental regulations are beginning to show signs of becoming overwhelmed by the shear mass of such information. The obvious response to this situation is for them to demand a prescribed degree of uniformity in all data that is submitted. Such *uniformity* is anticipated to be a means of ensuring a minimum required level of accuracy and reliability. Often these criteria primarily relate to the instrumentation used to make the final qualitative identification and quantitative measurement of the analytes being sought. In many trace analytical schemes (e.g., determination of chlorinated dibenzo-p-dioxins and dibenzofurans (CDDs/CDFs)) this approach to GLP is heavily dependent upon specifications related to the maintenance and operation of the gas chromatograph-mass spectrometer (GC-MS). Although these GC-MS criteria are indeed important to overall data quality, they are often not the primary reason for procedural variability and measurement inaccuracy. Since each analytical method that is designed to determine CDDs/CDFs has its own set of strengths and weaknesses, and these characteristics may vary when different analysts perform the method, it is virtually impossible to mandate GLP guidelines that yield unequivocal data and are economical to perform in a reasonable amount of time. Therefore, the analytical chemist should not blindly adhere to such mandated GLP in the belief that such adherence will produce quality analyses, other considerations may well prove to be more important to achieving the desired goal of optimum reliability and speed.

DISCUSSION

The purpose of this dissertation is to invoke a sense of responsibility within the trace analytical community concerning evaluation of their work from the perspective of overall accuracy and reliability. While GLP and QA/QC routines have always been a significant portion of analytical methodology dealing with trace level determinations, their positive function has often been regarded only in terms of data reliability. From this perspective, GLP is a necessary evil whose application serves to proportionally increase the analysis time for any given procedure. Perhaps this is true if mandated GLP routines require extra time to accomplish and do not effect the desired improvement in data quality. However, when appropriately applied, GLP procedures can be shown to reduce the analysis time necessary to derive data within a defined acceptable error.

The American Chemical Society has published *Principles of Environmental Analysis* to aid in the design and application of trace analytical methodology.[1] From information therein, it can be demonstrated that the relationship between the number of samples necessary to be examined in order to obtain data within a defined acceptable error is described by the equation:

$$N = \frac{(Z\sigma)^2}{E}$$

where N equals the number of measurements necessary, Z equals the standard normal variate based upon the level of confidence, σ equals the standard deviation of the method, and E equals the tolerable error in the estimate of the mean. If for a given analysis, constant values are assumed for Z and E, then $N \propto \sigma^2$. From this it can be seen that any changes in the method which increase σ will produce an exponentially larger increase in N, hence when a high degree of accuracy is required, an analytical method with a small σ will have a smaller value of N and therefore require fewer measurements. Considering that one of the primary purposes of GLP is to reduce σ, then by building into a given method the highest degree of reliability, it is possible to actually decrease the amount of time necessary to run N analyses. It is this line of reasoning which supports the appropriate implementation of GLP techniques to improve methodological reliability, namely, because such mechanisms actually result in significant times savings over the long term.

Perhaps the greatest problem faced by the analytical chemist today is deciding which GLP procedures are most important for reducing σ. Often governmentally mandated GLP routines address laboratory issues which are important from the perspective of demonstrating compliance with accepted policies concerning unit operations (*e.g.*, instrument maintenance log books, records of reference standard sources, balance calibrations, GC-MS operating conditions and detector criteria), however, while indeed important to a given method, such items are usually not the most significant source of problems which result in reduced reliability. We as analytical chemists recognize that for the most part, GLP associated with those aspects of our methods which produce unaccountably high or low results are the true culprits needing maximum attention. Unfortunately, we are not often asked to report on our GLP efforts in these areas.

UNACCOUNTABLY HIGH results shall be the first topic of concern. We have observed that contaminated standards or reagents, interferences in the sample itself, or matrix effects resulting from co-extractives in the sample extract are often primary sources of such positive excursions. We have taken extra precautions in the practice of our CDDs/CDFs methodology to minimize the potential for these problems to occur.

1) Considering isotopically-labeled internal standards, it should be recognized that they can be contaminated with native species (*e.g.*, native 2378-TCDD present in [$^{13}C_{12}$]-2378-TCDD) which may produce high results if unnoticed. Prior to the use of any labeled reference standards in a sample analysis, we perform an extensive series of GC-MS examinations to verify chemical identity, isotopic purity, chemical purity, and absence of analyte contamination. Interestingly, over the past ~10 years, ~40% of the labeled standards that we have purchased have failed to meet the simple criteria we have adopted for use in our methods.

2) Positive reagent blanks can originate from poor equipment cleaning practices, environmental contamination of sample extracts, or from contamination of chemicals or adsorbents used in the sample preparation. We have imposed a rigorous quality assurance program which involves analysis of a reagent blank with every set of samples processed simultaneously. A positive reagent blank can result in rejection of data and reanalysis of samples after the source of the contamination has been identified and/or eliminated.

3) Regardless of the resolution characteristics of the GC-MS instrumentation employed in a given method, interferences in sample extracts can be encountered. We have found that analyte identification criteria based solely upon GC-MS detection parameters, as often prescribed by governmentally mandated GLP routines, are not suitable to reliably preclude high results. Therefore, our approach couples such GC-MS criteria with additional parameters associated with chromatographic phenomena demonstrated by the specific analytes on their passage through the cleanup process. In essence, we define adsorbent isolation steps such that the retention characteristics of the analytes are specific enough to form an extended set of retention criteria which must be met to ensure accurate GC-MS identification and quantitation.

4) *Matrix effects* are a group of complicated and ill defined phenomena associated with sample extracts produced by nearly all cleanup methodologies. They are recognized by virtue of their influence upon GC-MS detector response characteristics. Typically such effects may cause high results by changing the response of the mass spectrometer and can often be observed by extraordinarily high *apparent* recoveries of labeled internal standards. While many analysts depend upon isotope dilution routines to correct for these phenomena, it is important to recognize that matrix effects can influence detector response differently for the various ion masses being monitored as well as vary from one region of the GC chromatogram to another.

UNACCOUNTABLY LOW results are a second major area of concern with regard to method reliability. Our experience indicates that incomplete analyte extraction, loss or degradation of the analyte during sample cleanup, or matrix effects are the most typical culprits. These problems can usually be reduced and/or eliminated by a careful evaluations of each step in the sample preparation.

1) Analyte extraction efficiency is a commonly overlooked and underestimated phenomenon. Too often analysts rely upon the recovery of labeled internal standards that are mechanically introduced into the sample to correct for apparent losses (or failure to extract) of native analytes. As long as the original sample matrix and the internal standards begin the workup as a solution such logic is reliable, however, in cases where this criteria is not met, such logic is faulty and may lead to low results. Considering the common CDDs/CDFs determination in solid matrices which are not amenable to dissolution, it is necessary to carefully evaluate both the extraction solvent and procedure to ensure that they are as exhaustive in character as possible.

2) Loss or degradation of the analytes can be controlled by monitoring recoveries through each step in the method. An understanding of the chemical and physical properties of the analytes will not only improve recoveries, but also, can allow optimization of each individual step. We extend such GLP criteria to chromatographic adsorbents in our methods because experience indicates that these materials are notorious for exhibiting changes in retention characteristics despite rigorous control of their manufacture and use. Typically we evaluate each adsorbent-based step to determine its capacity for analytes and the combination of analytes plus sample matrix constituents, thereby reducing the potential for *overloading*. In addition, we typically prepare adsorbents in small batches to reduce storage time before use, and also perform appropriate analyte specific optimizations and calibrations for each batch prepared.

3) We have typically observed that *matrix effects* responsible for the generation of low results are related to exceeding the capacity of the cleanup procedures for either the analytes, related compounds, or co-extractive species. Such *overloading* can change the chromatographic retention of the analytes. In general, these effects can be

minimized by decreasing the analysis aliquot or by using a high capacity pretreatment step to remove bulk matrix constituents. Whenever increased sample sizes are employed as a mechanism to achieve greater *sensitivity*, it is often prudent to prepare the final residue via passage of several smaller aliquots through the cleanup rather than as a single unit. Although such techniques seem counterproductive with regard to reducing analysis time, one should consider the adage that we never seem to have enough time to do it right the first time, but we always find time to do it over when something goes wrong.

CONCLUSIONS

The GLP routines described herein have been rigorously applied within the author's laboratory over the past ~13 years to govern and control the reliability of CDDs/CDFs methodology. As a demonstration of their cumulative effect, one can examine the analytical data produced from a recent collaborative study that was conducted to determine fortified levels of specific CDDs/CDFs in human adipose tissue at the 5 - 50 pg/g concentration level.[2] The goal of the study was to investigate *state of the art* analytical measurement capabilities in this matrix during the time period of ~1985, and as such, eight laboratories that were highly skilled in conducting these determinations were invited to participate. Referring to the data presented in Table 1, it can be seen that Lab #2 (the author's) avoided producing either unaccountably high or low results (determined by criteria agreed upon by each of the participants). In essence, this methodological ability was in large part responsible for a σ of 19% and an average analyte recovery of 97.7%. Taking an antagonistic view, it is to be noted that the average analysis time per sample for Lab #2 was also the greatest of all participants. From a short term perspective, this information can mean that Lab #2 was the slowest to generate analytical findings. However, remembering that $N \propto \sigma^2$, when the standard deviation of the recoveries for each laboratory is used in the equation:

$$\text{Relative time to N} = \frac{(\sigma_{\text{LAB \#}})^2 \left(\frac{\text{days}}{\text{sample}}\right)}{(\sigma_{\text{LAB \#2}})^2 \left(\frac{2.6 \text{ days}}{\text{sample}}\right)}$$

to determine the relative time to analyze N samples with respect to that obtained by Lab #2, we see from the results in Table 1 that Lab #2 can actually generate data with a specified reliability in the least amount of time.

Table 1. *Collaborative study data.*[2]

	Lab #1	Lab #2	Lab #3	Lab #4	Lab #5	Lab #6	Lab #7	Lab #8
Avg. Analysis Time (Days / Sample)	0.9	2.6	0.6	1.0	1.0	1.5	-----	2.5
Number of Values Unaccountably High	4	0	2	6	2	3	16	3
Number of Values Unaccountably Low	0	0	0	1	10	8	7	0
Avg. Recovery	115 %	97.7 %	114 %	147 %	97.0 %	50.3 %	135 %	151 %
σ (Std. Dev. of Recovery)	51 %	19 %	48 %	94 %	143 %	39 %	123 %	64 %
n (number of measurements)	26	23	23	26	23	26	23	25
Relative Time to N	2.5	1	1.5	9.2	22	2.4	-----	11

REFERENCES

1. L. H. Keith, W. B. Crummett, J. Deegan, R. A. Libby, J. Taylor and G. Wentler, Anal. Chem. 55, 2210–2218 (1983).
2. P. Albro *et al.*, Anal. Chem. 57, 2717–2725 (1985).