David Buckley & Associates Pty Ltd
(DBA) ABN 39 103 221 129
7 Catani Place Monash ACT 2904
Australia www.navigategmp.com
Phone/fax (61) 2 62914626
infor@navigategmp.com

Document #		Rev.:
BIO003:	**IND REV**	**1**

Title:

OOS - FDA and FEDERAL COURT INTERPRETATION OF GMP cGMP ON TEST FAILURE EVALUATIONS

The following is a U.S. federal court interpretation of CGMP requirements for addressing out-of-specification quality testing results. The discussion is contained in a February 4 ruling by Newark District Court Judge Alfred Wolin on the case of U.S. v. Barr Labs. Wolin's analysis is based on extensive testimony by several expert witnesses during the case hearings. Citations to the testimony are deleted in the excerpt below.

DRUG TESTING OVERVIEW

Testing lies at the heart of a drug manufacturer's successful operation.
Through testing, companies validate their processes and ensure the quality of batches for release. As the Forms [FDA] 483 suggest, much of this current litigation stems from allegedly defective testing practices. With the mechanics of test-taking left undefined by the regulations, before discussing the specific allegations against Barr, the Court will outline the CGMP-required parameters which will guide its evaluation.

FAILURES

In the government's view a batch failure occurs each time an individual test result does not meet the specifications outlined in the USP or the firm's ANDA. In contrast, Barr does not classify initial out-of-specification results as batch failures. Instead, only after confirming out-of-specification results with additional testing, pursuant to the firm's predetermined testing procedure, would Barr conclude that a batch failed.

Out-of-specification results obtained in the laboratory fall into three general categories:
(1) laboratory error;
(2) nonprocess-related or operator error; and
(3) process-related or manufacturing error.

Laboratory error can result from an analyst's mistake or malfunctioning laboratory equipment Examples of analyst error include mistakes in calculations, the use of incorrect standards for comparison, and simple mismeasurement. Those human and mechanical errors which occur during the manufacturing process cause nonprocess-related errors. For example, manufacturing equipment may malfunction or an operator may fail to add the proper amount of an active ingredient. In contrast, process-related problems, such as an incorrect mixing time, occur even though the workers and machines function properly.

While each type of problem is a matter of great concern which requires some form of corrective action, only non-process related and process-related errors properly are labelled failures. As [FDA Mid-Atlantic Region Investigator David] Mulligan acknowledged, all failures are not alike. An out-of-specification result identified as a laboratory error by a failure investigation or an outlier test, or overcome by retesting is not a failure. Thus, the Court is unwilling to adopt the government's view of failure.

FAILURE INVESTIGATIONS

Only with an investigation will a firm be able to identify the cause of an out-of-specification result. CGMP requires a thorough investigation following: "any unexplained discrepancy (including a percentage of theoretical yield exceeding the maximum or minimum percentages established in master production and control records) or the failure of a batch or any of its components to meet any of its specifications....[which] shall extend to other batches of the same drug product and other drug products that may have been associated with the specific failure or discrepancy. A written record of the investigation shall be made and shall include the conclusions and followup.

The government argues that an adequate failure investigation requires a timely, thorough, and well-documented review of the problem, which yields a written record containing:
(1) the reason for the investigation;
(2) a summation of process sequences that may have caused the problem;
(3) the corrective actions necessary to save the batch and prevent recurrence;
(4) a list of other batches and other products possibly affected along with their investigation results; and finally,
(5) comments and signatures of production and quality control personnel regarding approval of any material reprocessed after additional testing. Barr advocates a sliding-scale approach, claiming that the nature of the failure should govern the intensity of the investigation.

In accordance with the CGMP-required for-cause failure investigation, the goal of every such inspection is to determine into which of the three failure categories the problem falls. The degree of inquiry required successfully to complete this task may vary with the object under investigation. As a result, a full investigation of the type the Government outlined always will not be necessary.

The issue of failure investigations first arises when testing produces a single out-of-specification result. Before proceeding to retest, this unhappy occurrence must be met with a step-by-step review of the suspect laboratory tests. Specifically, the analyst who performed the test must report the problem to a supervisor, and the two technicians must conduct an informal laboratory inspection, reviewing the notebook which contained the out-of-specification result, discussing the testing procedure along with any required calculations and examining the instrument used. Thus the Court requires that a single out-of-specification result be met with more than "a laboratory investigation consisting principally of retesting."

Such an investigation, along with any conclusions reached, must be preserved with written documentation that enumerates each step of the review, in the form of a "computer generated flow sheet, or a check-list. This writing should be preserved in an investigation or failure report, and placed in a central file. In order to enhance this review process, each analyst conducting a test should follow a written procedure, checking off each step as it is completed.

Any easily identified analyst errors, such as calculation mistakes, should be specified with particularity and supported by evidence. In some instances, however, "the subtle influences which can result in test variability are not apparent when such an assay or test investigation is carried out." Thus because it can be difficult to pin down the exact cause of a problem, it is unrealistic to expect that the cause of analyst error always will be determined and documented.

In recognising the existence of less readily identifiable mistakes and the influence of variables unrelated to the purity or potency of the drug under scrutiny, the Court does not intend to create a means for firms to avoid the performance and documentation of an informal laboratory investigation. The inability to identify an error's cause with confidence affects retesting procedures, but does not affect the failure inquiry required for initial out-of-specification results.

Other problems more serious than single out-of-specification results, from multiple out-of-specification results to product mixups and contamination, require full-scale inquiries involving quality control and assurance personnel in addition to laboratory workers in order to identify the exact process or nonprocess-related errors.

Extending beyond the laboratory and often labelled formal investigations, these inquiries should follow the outline the Government provided, with firms paying particular attention to any necessary corrective action, whether reprimanding, retraining or firing employees, remixing batches or adjusting processes. Thus, in the failure report firms must:
(1) state the reason for the investigation;
(2) provide a summation of the process sequences that may have caused the problem;
(3) outline the corrective actions necessary to save the batch and prevent a similar recurrence;
(4) list other batches and products possibly affected, the results of their investigations, and any required corrective action; and finally,
(5) preserve the comments and signatures of all production and quality control personnel who conducted the investigation and approved any reprocessed material after additional testing.

The outcome of the failure investigation will determine whether additional batches of the same product and related products also require remedial measures. Process-related errors suggest the need to examine other batches of the problem product as well as other products made according to similar procedures. Addressing nonprocess-related errors requires an examination of other batches or products the trouble-making employee or machine may have handled.

Thus, the elements of a 'thorough" investigation necessarily will vary with the nature of the problem identified. However, all failure investigations must be performed promptly, within thirty business days of the problem's occurrence, and recorded in written investigation or failure reports.

OUTLIERS

The outlier test provides an alternative means of invalidating an initial out-of-specification result. If the failure investigation of an initial out-of-specification result proves inconclusive, firms searching for a better explanation can utilise this method.

Significant limits accompany the outlier test, however. The USP specifically warns against using outlier tests too often, and thus, as a general rule, 2 firms must be careful not to reject results frequently on this basis.

In addition, the utility of the outlier test varies with the type of assay performed. The USP expressly allows firms to apply this test to biological and antibiotic assays, but is silent on its use with chemical tests. Although some experts advocated use of the outlier method for chemical assays, other testimony suggested that firms generally do not rely on outlier tests to invalidate chemical test results. In the Court's view the silence of the USP with respect to chemical testing and outliers is prohibitory.

The substantial innate variability of microbiological assays supports this distinction. Chemical assays are considerably more precise than biological and microbiological assays, since only the latter testing is "subject to whims of microorganisms."

RETESTING

In addition to triggering a failure investigation, out-of-specification results also generate the need for retesting. A retest is defined as additional testing on the same sample, and thus it necessarily follows an initial test. An analyst performing a retest takes the second aliquot from either;
(1) the sample that was the source of the first aliquot; or
(2) the larger sample previously collected for laboratory purposes.

These procedures are equivalent. Thus, whether retesting is performed at the finished product or blend stage, such testing should be performed from the same bottle of tablets or capsules and the same drum or mixer, respectively.

Retesting is proper only after a failure investigation is underway, since the outcome of the failure investigation itself, in part, determines when retesting is appropriate. Retesting is necessary if a failure investigation indicates that analyst error caused an initial out-of-specification result. A retest is similarly acceptable when review of the analyst's work is inconclusive. In these instances, retesting substitutes for or supplements the original tests which have been rejected or questioned, respectively. In the case of nonprocess and process-related errors, however, retesting is suspect.' Because the initial tests are genuine, in these circumstances. additional testing alone cannot infuse the product with quality.

As a general matter, the amount of retesting required also varies with the problem identified. Out-of-specification results attributed to analyst error require limited retesting. Here, retesting merely supplants the first round of initial tests. More extensive retesting should follow an inconclusive failure investigation, since firms need to determine whether the out-of-specification result is a mere anomaly or a reason to reject the batch.

The USP contemplates retesting for quality control purposes, although it does not prescribe or recommend the number of individual tests that must be performed in order to reach a definitive conclusion about the quality of a product. Thus, the number of retests performed before a firm concludes that an unexplained out-of-specification result is invalid or that a product is unacceptable is a matter of scientific judgment. Yet the goal of retesting is clear; firms must do enough testing to isolate the out-of-specification result, in order to reach the point at which the additional testing overcomes the out-of-specification result.'

Nevertheless, retesting cannot continue ad infinitum. Because such a practice is not scientifically valid, a firm's predetermined testing procedure should contain a point at which testing ends and the product is evaluated. At this time, if the results are not satisfactory, the batch must be rejected.

When evaluating retest results, it is important to consider them in the context of the overall record of the product. Relevant to this review are the history of the product,' the type of tests performed, and any results obtained for the batch at other stages of testing." As such, retesting determinations will vary on a case-by-case basis, a necessary corollary of which is that an inflexible retesting rule, designed to be applied in every circumstance, is inappropriate.

RESAMPLING

Resampling, in contrast, is a more controversial practice. Typically resampling occurs after the initial test and the retests have produced out-of-specification results, thereby indicating a more serious problem. When performing a resample, an analyst leaves the laboratory and takes a new sample from the universe of the batch.

Resampling is appropriate where provided by the USP, as in cases of content uniformity and dissolution testing. Similarly, in the limited circumstances in which a failure investigation suggests that the original sample is unrepresentative, resampling is acceptable. Evidence, not mere suspicion, must support a resample designed to rule out preparation error in the first sample.'

Absent these limited exceptions outlined above, however, firms cannot rely on resampling to release a product that has failed testing and retesting.

REMIXING

The need for remixing arises during the blend stage when testing reveals problems with content uniformity. The regulations themselves allow reworking, which essentially is remixing. As evidenced by the Generic Drug Office directive and the consent agreement between the FDA and Eli Lilly, remixing is allowed in this circumstance.

The instance of remixing, however, is directly related to its propriety. Occasional remixing is acceptable, but frequent or wholesale remixing is not. The need to remix often provides a clear indication that the process is invalid, and casts doubt on those batches that passed through testing without incident.

AVERAGING

Although averaging test data can be a rational and valid approach, as a general rule, firms should avoid this practice, because averages hide the variability among individual test results.

This phenomenon is particularly troubling if testing generates both out-of-specification and passing individual results which when averaged are within specification. Here, relying on the average figure without examining and explaining the individual out-of-specification results is highly misleading and unacceptable.

Although averaging camouflages variability, an average may provide more information about the batch's true assay value than any single test result. Thus, in the case of microbiological assay testing, the USP prefers an average when reaching an ultimate judgment about the quality of the product. It is good practice to include out-of-specification results in the average, unless an outlier test indicates that an out-of-specification result is an anomaly.

Finally, the average of individual content uniformity tests at the finished product stage can act as a proxy for the assay value. Though this estimate cannot substitute for final product assay testing, it can provide some information about a batch.

RELEASING A BATCH FOR DISTRIBUTION

Section 211.165(a), in relevant part, provides: 'For each batch of drug product, there shall be appropriate laboratory determination of satisfactory conformance to final specifications for the drug product, including the identity and strength of each active ingredient prior to release." And section 211.165(f) specifies: "Drug products failing to meet established standards or specifications and any other relevant quality control criteria shall be rejected. Reprocessing may be performed. Prior to acceptance and use, reprocessed material must meet appropriate standards, specifications, and any other relevant criteria."

The USP provides the "established standards" to which the CGMP regulations refer and upon which firms rely to release their drug products for distribution to the public. These specifications are absolute.

Under section 211.165(f), the government argues, a single out-of-specification result that cannot be invalidated defeats a batch. Barr reads the statue more liberally and refuses to equate a failure to meet specifications with obtaining one out-of-specification result from many quality performance measurements.

Similarly, because the USP specifies the standards for release testing, the government contends that section 211.165(f) removes any uncertainty that might require the exercise of scientific judgment. The absolute approach the government recommends, under which a batch fails if one tablet out of one million tablets tested produces an out-of-specification result, is extreme. Instead, in the Court's view, section 211.165, through the "appropriate laboratory determination" language and the allowance of reprocessing, suggests that scientific judgment can play a role when firms decide whether to release a batch to the public.

In light of the case-by-case variability scientific judgment introduces, the Court cannot articulate specific procedures for release decision-making. Instead, in this context, firms should follow the general retesting guidelines. However, firms cannot justify release when only fifty or sixty-six percent of the finished product tests for a particular batch produce passing results. For example, with an assay limit of 90 to 110, test results of 89 and 90, or 89, 91, and 91, or two 89s and two 92s all should be followed by more testing. The amount of additional tests required would be a matter of scientific judgment, as informed by other relevant data. The goal is to distinguish between an anomaly and a reason to reject the batch.

While experts disagree about the relative importance of finished product and blend testing, it is clear that the release evaluation depends, in part, on the background of the batch and product. Secondary parameters, such as physical properties, blend evaluations, time of mix, weight, thickness, and friability, affect the actual finished product results as well as their reliability. The lesson for firms and the FDA is that context and history inform many final conclusions.

BLEND TESTING

An important aspect of drug manufacturing, blend testing gives firms an opportunity to discover and remedy in-process problems before batches reach the final stages of production. Because finished product testing is limited, blend testing is necessary to increase the likelihood of detecting inferior batches.

(a) SAMPLE SIZE

An element of blend testing which influences the ultimate test results is sample size. No regulation, guideline, or publication requires any specific blend sample size. The CGMP regulations merely provide that 'in-process materials shall be tested for identity, strength, quality and purity as appropriate.'

In accordance with the preamble to the CGMP regulations, the Court must construe the "as appropriate- phrase to permit 'reasonable, albeit variable interpretations.- The testing procedures a firm chooses, however, must be logical and effectual. Paraphrasing Inspector Mulligan, the Court must ask whether 'what they are saying makes sense.'

Implicit in the 'as appropriate language, the government argues, is a sample size requirement equal to one to three times the product's run weight. Driving this theory is the concern that a larger sample will dilute or even negate any non uniform in the blend. Barr argues that a variety of sample sizes comply with CGMP and that the final choice is a matter of scientific judgment.

Although sample size is a question of scientific judgment, the sample chosen must advance the purpose of the test. Thus, what is "appropriate" may vary with the type of test performed.

Content uniformity testing, designed to detect the adequacy of the mix by measuring variations in the potency of the blend, should be conducted with a sample that resembles the dosage size. Any other practice likely would blur differences in portions of the blend, and defeat the object of the test.

The Court appreciates the difficulty companies experience taking minute samples from large-volume blends. Indeed, testimony revealed that the smallest thief available can retrieve a 250-milligram sample, so in some cases firms cannot obtain a single-run-weight sample. As such, the Court will follow Dr. [Robertl Gerraughty's testimony and hold that the appropriate sample for content uniformity testing, in both validation and ordinary production batches, is three times the active ingredient dosage size.

In addition to assuring a more accurate measure of uniformity, this rule accommodates the need for retesting. In order to conduct an initial test and two retests, a standard testing practice in the industry, analysts need a three-run-weight sample. Under Inspector Mulligan's one-run-weight rule, in order to retest the same sample, firms must take additional samples from the same spot in the blend. Such a requirement would be onerous.

Implicit in the sample size determination for content uniformity testing is a prohibition on compositing multiple individual samples taken from different areas of the blend. Again, in order to detect uniformity problems, firms must avoid this practice which would conceal variations in the blend.

In contrast, blend assay or potency testing, designed to measure the strength of the blend, can accommodate larger samples. Although averaging the differences in the mixture, here a larger sample also provides a better indication of the overall percentage of active ingredient in the blend. In fact, a blend assay test conducted with a larger sample will be more representative of the final assay. Similarly, as variation detection is not the object of assay testing, the Court does not object to compositing."

(b) SITE OF SAMPLING

Also important in content uniformiy testing are the number of samples taken and how representative they are of the mix. The government again cites the 'as appropriate' language of section 211.110(c) to support its view that blend content uniformity testing should be conducted with samples taken from the mixer and not the drum. Barr maintains that firms are free to sample from either the mixer or the drum and, further, that in some cases, testing from the drum is preferable. Once again, no regulation, guideline, or publication expresses a preference for blend testing in the mixer or the drum.

Expert testimony revealed that firms test from both the drum and the mixer and that either practice is acceptable under CGMP. Thus, the Court is not prepared to prescribe a particular location for blend testing.

Rather, the factor of more concern to the experts, and thus the Court, is the representativeness of the sampling technique. To detect poor uniformity, firms must take samples from 'places that might be problems," such as weak and hot spots in the blend. Indeed, Barr conceded that Inspector Mulligan's concern about in-process blend testing in weak spots of the mixer is valid at the process validation stage. Thus, whether sampling from the mixer or the drum, firms must demonstrate through validation that their technique is representative of all portions and concentrations of the blend.