



Identification of Out-of-Trend Stability Results, Part II PhRMA CMC Statistics, Stability Expert Teams

Oct 2, 2005

By: PhRMA CMC Statistics, Stability Expert Teams
Pharmaceutical Technology

High-quality stability data are critical to the pharmaceutical and biopharmaceutical industries. These data form the basis for justifying specification limits (also called *acceptance criteria* in the International Conference on Harmonization [ICH] Q1 and Q6 guidances), for setting and extending product expiration dates, and for establishing product label storage statements. Annual, routine stability studies also can be used to support product or process modifications and are vital for ensuring the continuous quality of production batches. To facilitate the prompt identification of potential issues, and to ensure data quality, it is advantageous to use objective (often statistical) methods that detect potential out-of-trend (OOT) stability data quickly.

In general, OOT stability data can be described as a result or sequence of results that are within specification limits but are unexpected, given the typical analytical and sampling variation and a measured characteristic's normal change over time (e.g., an increase in degradation product on stability).

OOT stability results may have a significant business and regulatory impact. For example, if an OOT result occurs in the annual production batch placed on stability, the entire year's production could be affected, depending on the OOT alert level. Systems that are useful for detecting OOT results provide easy access to earlier results so that they can track stability data and trigger an alert when a potential OOT situation arises. Procedures describing OOT limits, responsibilities for all aspects of an OOT system, and investigational directions are all components of a company's quality system.

Representatives from the US Food and Drug Administration and 77 individuals from more than 30 Pharmaceutical Research and Manufacturers of America (PhRMA) member companies met in October 2003 to address important stability OOT alert issues. Their discussions underscored the need to track OOT stability data, associated issues, and potential statistical or other approaches to establish OOT alert limits. Breakout sessions captured participants' perspectives on FDA expectations and guidances, acceptable OOT identification methodologies, and implementation challenges.

Workshop attendees concluded that the main focus for OOT identification and investigation should be annual, routine production stability studies rather than primary new drug application (NDA) batches because historical data are usually needed to determine appropriate OOT alert limits.

In addition, attendees clearly expressed a desire for a simple approach to tracking stability data that would allow at-the-bench, real-time detection of OOT results. This capability is important from an implementation standpoint so that a timely investigation can be initiated. Timeliness is especially important when an analytical error is suspected of causing OOT results. It was recognized that an efficient approach with such properties might be difficult to achieve.

Another recommendation was the need for standard operating procedures (SOPs) that address how to handle (e.g., identify, investigate, conclude, take action) OOT data. Workshop participants believed that a separate FDA guidance is not needed to develop such SOPs because the existing out-of-specification (OOS) guidance should suffice.

Three types of OOT results—which determine the degree and scope of an OOT investigation—were identified at the workshop. This article expounds on these outcomes and addresses other questions raised during the workshop and in a previous article (1). Recommendations put forward in this article are not requirements, but rather, are possible alternatives for handling OOT data.

Types of OOT data

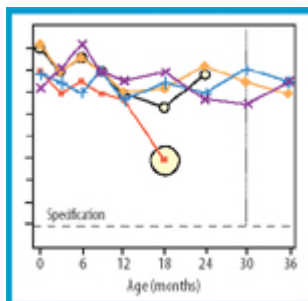
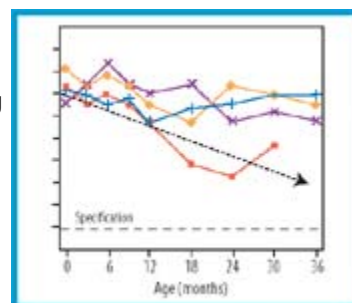


Figure 1: An analytical alert (theoretical data, see endnote).

A *process control alert* occurs when a succession of data points shows an atypical pattern that was possibly caused by changes to the laboratory or manufacturing process. These data points might originate from the same stability study (see Figure 2) or from multiple studies assayed within a

OOT alerts can be classified into three categories to help identify the appropriate depth for an investigation. In this article, OOT stability alerts are referred to as *analytical*, *process control*, and *compliance* alerts, in order of implied severity. As the alert level increases from analytical to process control to compliance alert, the depth of investigation should increase.

Historical data are needed to identify OOT alerts. An *analytical alert* is observed when a single result is aberrant but within specification limits (i.e., outside normal analytical or sampling variation and normal change over time) (see Figure 1).



reasonably close timeframe (within a few weeks) (see Figure 3). As these Figures 2 and 3 indicate, the trends for the batches in question vary substantially from those of comparable batches. Despite the deviating trends, no pending potential OOS situation occurs.

Lastly, a *compliance alert* defines a case in which an OOT result suggests the potential or likelihood for OOS results to occur before the expiration date within the same stability study (or for other studies) on the same product (see Figures 4 and 5).

Figure 2: A process control alert (theoretical data, see endnote). Results from multiple time points for a study do not follow the same trend over the 36-month shelf-life expiration as do other studies. The study is not expected to fall below the specification limit through the 36-month expiration.

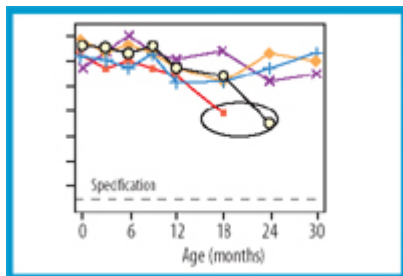


Figure 3: A process control alert (theoretical data, see endnote). Within the past few months, results from two studies are unexpectedly low, each initiating an analytical alert. Neither study is expected to fall short of the specification limit through the 30-month expiration.

Similarly, tests with multiple stages can be classified into the three OOT alert levels, but this task is much more challenging and less straightforward. For example, consider the three-stage *USP* dissolution test. If Stage 2 testing is rarely required, then an analytical alert for *USP* dissolution testing might be a single, unusual minimum or average dissolution. If the product usually passes Stage 1 and a new stability study goes to Stage 2 several times, then this result might signal a process control alert. If the study's dissolution average shows a decreasing trend with time that indicates difficulty remaining above the registered Q-value through expiration, this fact could indicate a compliance alert.

Very minimal guidance was provided on this topic at the workshop, though some attendees believed it is an area that lacks clarity and consistency.

SOP requirements

The workshop attendees suggested that companies develop SOPs for the review and investigation of OOT stability results. FDA participants pointed out that current good manufacturing practices (CGMPs) require a written stability-testing program to assess the stability characteristics of drug products (211.166) and require written procedures for conducting a thorough investigation of any unexplained discrepancy (211.192). Moreover, CGMPs require written procedures for evaluating the quality standards of each drug product at least annually (211.180). A review of the OOT alert procedures' performance might coincide with the annual product review. When preparing an SOP for handling OOT test results, it is prudent to consider risks to product quality when determining the company's assessment of OOT test results. The depth of an investigation and the corrective measures taken may depend on the potential or implied risk to product quality.

The following five SOP topics are areas that companies may want to address. These points were not discussed by conference participants.

How to set alert limits. SOPs may identify scientifically sound approaches to determine procedures with associated limits for the identification of each type of OOT alert for stability test attributes such as assay and degradation products. Once an approach is identified, some potential topics to be included in the SOP are data requirements, personnel responsible for setting the limits, and methods used to detect each type of OOT alert.

For new products, available stability data will be limited. Established products often will have sufficient historical data. The methods for detecting OOT results may vary depending on the type of OOT alert. Ideally, a method for each OOT type should be specified.

How to use limits. The SOP may contain a section about how OOT alert limits are applied and which personnel will be responsible for comparing new stability results with the OOT alert criteria and timeliness of the review.

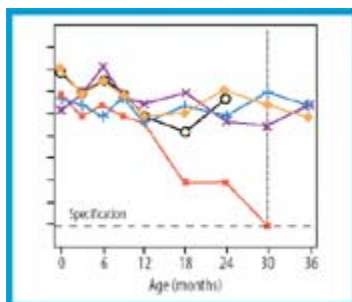


Figure 5: A compliance alert (theoretical data, see endnote). After two analytical

Investigation. Investigation requirements could include personnel responsibilities with defined timelines, documentation requirements, and proper internal notification requirements. The investigation's depth required for each OOT alert will depend on potential risks to product quality. The SOP should not be overly prescriptive because the steps taken to investigate OOT results will depend on the nature of the initial finding and the investigation findings. Nonetheless, SOPs should provide basic criteria for conducting the investigations. Moreover, the steps taken to investigate OOT results should be scientifically defensible, and the rationale for the investigation's depth should be documented. Although an OOT result usually will be investigated as soon as possible after it is identified, it also can be addressed more broadly during the annual product review. The notification requirements and the department

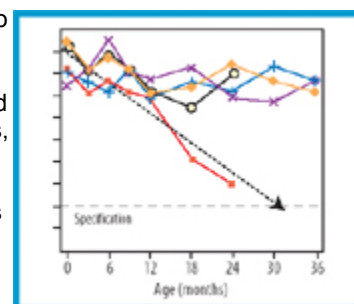


Figure 4: A compliance alert (theoretical data, see endnote). Results from multiple time points for a study do not follow the same trend as other studies. The study is at risk of falling below the specification limit before the 36-month shelf-life expiration.

alert results in a study, a result is observed right on the specification limit at the 30-month expiration time point. Because this study may represent many other batches, it is a potential compliance concern.

responsible for performing the annual product reviews also could be included.

Limits for truncated data. The SOP may specifically address truncated data such as for degradation products, whereas the approaches discussed in this article may not be appropriate because of the truncation. An example of truncated data is unreported data that are lower than the ICH reporting threshold. Degradant results may be higher or lower than the ICH reporting threshold to time point. In addition, results may be reportable only at later time points as the levels increase depending on the amount of truncation (ICH reporting threshold). Setting data-driven OOT alert limits in such situations may not be possible. To facilitate the quantitative evaluation of the data, it is suggested that degradation product and impurity results be available to at least two decimal places.

Periodic review of limits. SOPs may specify a periodic review process for OOT alert limits. The frequency of the review and the personnel responsible for completing it may be addressed in the SOPs. This assessment can involve identifying a previous result that was once within trend but is now an OOT result. Or, it may find that a previous OOT result is now within trend. It is natural that earlier conclusions be revised in light of new information because developing OOT alert criteria is an active and continuing process.

Depth of investigation

It is important to point out that by itself, an alert provided by a method designed to detect unexpected events or degradation patterns is no proof that the batch in question has a deviating trend. The alert is only a signal to begin an investigation. The alert is not intended to make any conclusions. A conclusion can only be reached after the investigation of OOT stability data that follows.

An investigation's depth may depend on the nature of the data, the product history and characteristics being measured, the potential risk to product quality, and the potentially related safety or efficacy implications. The CGMP regulations in *CFR* 211.166 require an assessment of drug stability, whereas 21 *CFR* 211.192 requires an investigation of unexplained discrepancies by the quality control unit.

If an analytical alert is observed, a laboratory investigation may be performed. If the laboratory investigation is inconclusive, then the supervisor may take no further action at that time and monitor subsequent time points closely. A review of results for related or correlated test attributes also can facilitate the investigation. Depending on the product, history, and nature of the analytical alert data, the company may decide to investigate whether a manufacturing error occurred in cases in which a laboratory error is not conclusive. More data (*i.e.*, the next stability test point) may be needed to indicate whether the result is part of a long-term trend (process control alert) or just an isolated result. If subsequent results within the study and across other studies are not out of trend, then the initial analytical alert result may be an isolated incident that requires no further investigation.

A process control alert may indicate unexpected changes in product or assay performance. A stability study with an unusual trend may reveal that a characteristic's stability profile changed, whereas multiple analytical alerts may suggest that the measurement process is no longer in control. When a process control alert is evident, the investigation will often begin with an assessment of the effect of potential changes to the laboratory process (*e.g.*, changes in instrument, column, integration technique, or standard). It also may extend to the manufacturing process (*e.g.*, looking for changes in materials, personnel, equipment, and procedures).

Because a compliance alert identifies a situation in which a particular study (or several related studies) may not meet specification limits during the expiration period, one may need to conduct a thorough investigation. In general, this investigation begins with the laboratory process and expands to the manufacturing process if no conclusive laboratory root cause is identified. The manufacturing investigation might include other batches (from the same or related products) that may be associated with this predicted failure to identify whether the discrepancy is isolated or systematic. The investigation also could assess whether more analytical testing, an investigation into the manufacturing process, a product recall, or shorter dating is required. An investigation of a compliance alert aims for the early detection of potentially failing product and the identification of possible causes. If a root cause is determined, then appropriate action should be taken such as the identification of potential preventative actions.

Methods for detecting OOT results

Several statistical methods can be used to detect the three types of OOT alerts. Some simple methods for detecting analytical alerts were addressed at the workshop and are discussed in this article. These methods tend to require less statistical support and do not require a laboratory information management system (LIMS). The pros and cons for each analytical alert method also are discussed in this article. Other methods that can be used to detect potential process control and compliance alerts are only briefly discussed.

In addition, several important data considerations for planning an OOT program will be addressed. It is preferable to detect all three types of OOT results early, but process control and compliance alerts are more difficult to detect during product testing and are particularly difficult to detect for early time points in the study.

Detecting an analytical alert. It is important to discover analytical alerts as close to the result collection time as possible because they are primarily designed to detect laboratory errors. This process will maximize the analyst's chances to either confirm the result or to identify the cause of the aberrant result and correct it. Without having the ability to conduct the evaluation automatically within a LIMS, simple tools are desired.

The detection methods discussed in this article for analytical results involve relatively simple calculations. Other approaches (discussed in Reference 1) may be more effective if their implementation is feasible. In general, simple methods are easier to implement, do not require continuous statistical expert support, and can be used without a complex LIMS. They are most effective at detecting individual aberrant results (analytical alerts) and are not as good at detecting process control and compliance alerts.

Four such methods are:

- change from previous;
- change from previous (per unit of time);
- change from initial;
- observed value.

These methods are based on an initial review of historical data to generate the needed limit (important considerations when choosing the historical data set are described later in this article). The methods discussed are superior to the use of an arbitrarily fixed set of limits (*i.e.*, the one-size-fits-all approach) for all products and tests because the one-size-fits-all limits are invariably too narrow for some cases (causing over flagging or false alarms) and too wide for others (causing under flagging and failing to signal true OOT situations).

The calculations required to use these methods are:

- *change from previous*, which uses a change from the previous stability time point result;
- *change from previous (per unit of time)*, which also uses the change from the previous result, but standardized as the change per unit of time by dividing by the number of time units between results. For example, if the unit of time were months, and there were six months between results, the change from the previous would be divided by 6 to get change per month.
- *change from initial*, which uses the change from the initial (time zero) result;
- *observed value method*, which uses the observed values without any required manipulation of the data.

For each method, examples of the calculations and the establishment of the OOT alert limits are provided in the appendix of this article.

Once OOT alert limits are established, the observed value of the change (or simply the obtained result for the observed value method) is compared with the preset OOT alert limits. If the observed change is within the alert limits, no alarm is generated. If the change is outside the alert limits, it is considered a potential OOT result. The limits should be calculated from the historical stability data for the product and test item being evaluated. Information from a histogram or frequency distribution of these differences can be used when setting the OOT limits.

If the stability for these batches decreases over time, then the average change will be negative and the alert limits will tend to include mostly negative values. On the other hand, if the batches are stable, then the average change will be approximately zero and the alert limits will tend to include both positive and negative values. Because analytical alerts are intended to catch possible laboratory errors, setting both upper and lower limits is recommended.

For example, the OOT limits can be defined by the equation,

$$\text{average change} \pm (K \times \text{standard deviation of the change})$$

in which K is either a multiplier such as $K = 3$ (if data can be assumed to be normally distributed) or the t -statistic ($df, 1 - \alpha$), in which df represents the degrees of freedom and α is the risk level. A thorough discussion of α is presented in the appendix of this article. Alternatively, the limits can be defined as the lower and upper observed α and $1 - \alpha$ percentiles of this distribution of differences. This distribution-free approach is preferable for cases in which data are not expected to be normal.

Caution is needed when deciding which value of α or K to use. The appendix offers points to consider when setting the α level. Similar issues apply to setting the value of K , except that there is an inverse relationship between α and K . When α is smaller, K should be larger, and *vice versa*.

Pros and cons of various methods to detect analytical alerts. *Stable products.* The following are some special notes about using detection methods on stable products.

- If the batches are expected to be stable, either the *change from previous* or the *observed value* method could be used. Although the *observed value* method is easier to calculate, it assumes that all batches begin at the same starting value. Initial differences among starting values will be included in the variability. Hence, this approach may not be as sensitive as other methods for catching analytical alert results. If variations in results are caused largely by assay variability (and not true batch-to-batch differences or true degradation), then the *observed value* method may be ideal.
- Occasionally the *change from previous* method tends to identify an OOT result too late. For example, the method detects the jump back to expected levels rather than the drop from expected levels. This method may also generate flags both on the drop and when the batch returns to normal levels.

Products with change over time. Special notes about using detection methods on products that change over time are as follows:

- If the batches are expected to change over time (*i.e.*, they are not stable), then either the *change from previous (per unit of time)* or the *change from initial* method is appropriate. The choice depends on how large the expected change over time is relative to the assay variability.
- The *change from initial* approach usually catches an OOT result sooner than other techniques. Nonetheless, one disadvantage of the approach is that it depends heavily on the initial result because all differences are compared with the initial result for that batch. Separate limits also must be generated for each of the stability time points; one set of limits would not suffice.
- If the change over time is nonlinear and the *change from previous (per unit of time)* method is used, then a transformation must be performed to linearize the data before the OOT limits are calculated. The *change from initial* method does not require any transformation for nonlinear data.
- The *change from previous (per unit of time)* method has a similar problem as the *change from previous* method discussed in the stable products section. This method can occasionally catch the jump back rather than the original drop.

Detecting a process control alert. The main focus of statistically based methods for detecting a process control alert is to determine whether the stability trend for the suspected batch is different from the expected stability trend (see Figure 2). The detection of a process control alert requires a review of the product's historical data. One needs a sufficient number of batches that have the expected stability trend to determine whether the suspect batch or batches could generate a process control alert. Because the methods for detecting a process control alert tend to be more complicated than those used to detect an analytical alert, they are also more difficult to implement at the time of result entry. It is possible that a process control alert could most easily be detected during an annual product review when the historical data for the product is available and can be evaluated.

The slope control chart is one method for detecting a process control alert (1). Another approach is to combine the suspect study with the historical studies and test for the poolability of slopes (see reference 2 for how to test for poolability). If the historical studies were poolable without the suspect study and are not poolable after including the suspect batch, then this study could be classified as a process control alert. Process control alerts that involve multiple studies (such as Figure 3) may be detected visually during the investigation of an analytical alert or during the annual product review.

Other potential methods can detect process control alerts (3). More theoretical work and investigation is needed in this area to determine the best statistical methods for various situations and data requirements to successfully assess a process control alert.

Detecting a compliance alert. The key difference in detecting a compliance alert *versus* detecting a process control alert is determining the potential of the study to fail before expiration. Detecting a compliance alert also requires reviewing historical data to determine the expected stability trend. As with the process control alert, this alert could be investigated during the annual product review. Nonetheless, it is important to detect a compliance alert as early as possible because the study in question may be at risk of failing shelf-life specification limits.

The methods mentioned for detecting a process control alert also can be used for detecting a compliance alert. But, the evaluation also must include an element for determining whether the study is expected to fail before expiration. One specific method for detecting a compliance alert is to fit the stability model (often a linear regression) to the suspect study and predict the result at the time of expiration. The prediction is compared with the shelf-life specification limits to determine the risk of the batch failing before expiration.

The statistical issues that must be explored further for a compliance alert include:

- which statistical methods are best and under what conditions;
- what are the minimum data requirements for the historical data and for the batch under investigation;
- how best to assess the risk of failing before expiration.

Historical data considerations for all alerts. Any statistical approach to detect a potential analytical, process control, or compliance alert includes a comparison with a limit that defines what is unexpected. This OOT alert limit is related to historical experience (*i.e.*, it is data-driven and calculated from a historical database). The choice of which data are appropriate to include in this historical database is critical; it is probably as important as the choice of statistical method. To avoid over flagging, one might set a minimum number of time points and a duration of available stability data for a given batch before applying this rule.

The historical database represents the history, whereas the OOT alert limit will be used in the future. Hence, it is necessary that the database also represent the future. If this is not the case, then one could establish too wide, too narrow, or wrongly centered OOT alert limits. This can lead to alarm failures (*e.g.*, missing important signals), an elevated frequency of false alarms, or a mixture of both undesired events.

One reason to limit the full historical database could be that some product or process improvement has been introduced, thus resulting in better stability characteristics of batches manufactured after the change. The improved stability can, for example, be caused by a change of a raw material supplier. Using all stability batches in such a situation would lead to OOT alert limits that are too wide. The appropriate action is to exclude batches manufactured before the change from the data set on which OOT alert limits are based. It is often helpful to plot historical results *versus* manufacturing order when judging which data subset best represents future batches. Nonetheless, care must be taken not to reduce the historical data set too much. A well-documented change control system facilitates these decisions.

A common concern is that very limited historical data are available, in particular immediately after product approval. One way to expand the database is to pool data, provided it can be shown that the degradation rates are not different. Pooling various product presentations (*e.g.*, other package sizes, strengths, container closures, and container orientations) can be considered. Such pooling also is beneficial because common OOT alert limits for various presentations usually are desired for practical reasons. Nonetheless, the suitability of pooling must be carefully evaluated because pooling two presentations with different stability characteristics might lead to inappropriate limits for both presentations. Statistical comparisons and standard scientific considerations might be helpful in this evaluation. Different degrees of pooling might be suitable for different test attributes.

The existence of possible aberrant individual results in the historical database is another potential concern. Retaining a single suspect result in the database can have a significant effect on the OOT alert limits. Therefore, it is advisable to screen the database for such results, review the results individually, and make an informed decision to exclude or retain them. Several statistical methods for identifying influential results are available. Excluding a few historical influential results should be of no regulatory concern because it would lead to tighter limits. Exclusions should not be too extensive because they will increase the frequency of false alarms.

In addition, a problem may occur when evaluating impurity or degradation product results. In many situations, data censoring and over-rounding (*via* the impact of the ICH reporting threshold or the limit of quantitation) make it very difficult to identify OOT results. Thus, the total daily intake of the degradation product (defining the ICH qualification threshold) and the specific levels of the impurity or degradation product relative to the various reporting thresholds must be taken into account when developing an approach to identify OOT results that warrant investigation.

As more data become available, one has an opportunity to update and fine-tune the OOT limits. In addition, the manufacturing process might be further optimized or improved analytical methods may be introduced. Such changes highlight the need to reevaluate whether the historical database (and thus the effective OOT alert limits) is suitable for future batches. Therefore, it is recommended to review the suitability of the OOT alert limits regularly and update them, if needed. This reassessment can be accomplished most conveniently during the annual product review.

Conclusion

To enhance stability data quality and to continuously screen for signs of potentially deviating product, methods are needed to identify and to deal with potential out-of-trend (OOT) stability results. At the October 2003 workshop, industry and the US Food and Drug Administration attendees identified the need to develop suitable methods for detection. This article presents some outcomes from the workshop and discusses related questions.

Three types of OOT alerts—analytical, process control, and compliance alert—are proposed based on the number and patterns of data points that are out of trend and the type of the resulting trend. Detecting each alert requires various statistical tools. The analytical alert methods are easiest to implement while the process control and compliance alerts tend to be more complex.

The depth of the OOT stability data investigation depends on the type of trend, the nature of the data, product history and characteristics being measured, risk-based assessment, and the potentially related safety and efficacy implications. In all cases, the identification of OOT alerts should be based on historical experiences.

The issues surrounding OOT stability results encourage individual pharmaceutical firms to create standard operating procedures (SOPs) for how to handle such situations. Among other features, SOPs may describe methods to define the limits for each type of OOT alert, steps to verify whether an alert is detected, and actions to be taken.

Acknowledgments

The following members of the PhRMA CMC Statistics and Stability Expert Teams collaborated on this article: Mark Alasandro, Merck & Company, Inc.; Laura Foust, Eli Lilly and Company; Mary Ann Gorko, AstraZeneca Pharmaceuticals; Jeffrey Hofer, Eli Lilly and Company; Lori Pfahler, Merck & Company, Inc.; Dennis Sandell, AstraZeneca AB; Peter Scheirer, formerly of Bayer Healthcare; Andrew Sopirak, formerly of AstraZeneca Pharmaceuticals; Paul Tsang, Amgen Inc.; and Merlin Utter, Wyeth Pharmaceuticals.

Authors from the PhRMA CMC Statistics and Stability Expert Teams are listed in the Acknowledgments section of this article. All correspondence should be addressed to **Mary Ann Gorko**, principal statistician at AstraZeneca, 1800 Concord Pike, PO Box 15437, Wilmington, DE 19850-5437, tel. 302.886.5883, fax 302.886.5155, maryann.gorko@astrazeneca.com

Submitted: Oct. 15, 2004. Accepted: Feb. 1, 2005.

References

1. Pharmaceutical Research and Manufacturers of America, RMA CMC Statistics and Stability Expert Teams, "Identification of Out-of-Trend Stability Results," *Pharm. Technol.* **27** (4), 38–52 (2003).
2. US Food and Drug Administration, *Guidance for Industry, Q1E Evaluation of Stability Data* (Rockville, MD, 2004).
3. W.R. Fairweather *et al.*, "Monitoring the Stability of Human Vaccines," *J. Biopharm. Stat.* **13** (3), 395–414 (2003).

Endnotes

* All figures show historical data and a stability study through the current time point. The plots include simulated data to demonstrate possible scenarios. The symbols represent various lots of simulated data. The Y-axis could be any measurement made on stability.

Appendix

Choice of a level. Given a specific statistical approach and a set of data, the out of trend (OOT) alert limit follows as soon as the α level (Type I error) is selected. The Type I error is the risk of getting a false alarm (*i.e.*, a signal to start an investigation although there is nothing unusual about the finding).

At first, it seems rational to select a very low α level. The drawback with this approach is that the chance to identify a truly aberrant result (the power) also is reduced, which is not desirable. The best way to have a low Type I error and high power is to use an efficient statistical evaluation technique. Such techniques, however, are often too complex for the applications discussed in this article.

Many considerations must be balanced when selecting an α level. For example, some test attributes might be considered more critical for the product, and therefore a higher α level may be appropriate.

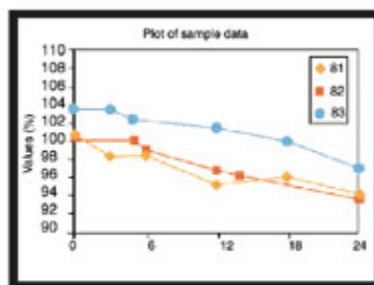


Figure 6: A plot of sample data.

Figure 6: A plot of sample data.

The issues associated with limited data can be balanced by selecting a low α level initially, and increasing the level as more data becomes available, unless the lower number of degrees of freedom ($n - 1$) associated with the t -statistic provides a higher multiplier without the need to adjust the α level. One also must consider the depth of investigation following an alarm; if the follow-up investigation typically is very complex, a lower α level may be suitable.

In summary, it is not possible to specify in general what is the appropriate α level. This level must be decided on a case-by-case basis depending on the unique product characteristics, test item, amount of data available, and the expected level and complexity of the investigation following an alarm. An assessment of what values are and are not caught by various methods when applied to historical data is encouraged; this evaluation will aid in the determination of what a level should be used.

It is important to remember that whatever a level is chosen, this number represents the frequency of alarms that

will lead to inconclusive investigations. Several investigations without conclusive cause will thus be expected and natural.

Calculations of OOT alert limits using various methods intended to detect an analytical alert. The following description is an example of how to calculate OOT alert limits for each analytical alert detection method discussed in this article. For simplicity, the same example is used to show the calculations for all four methods. The methods shown are not necessarily the best method for this data set. The purpose of the example is only to show the calculations. The examples are only illustrative in nature, with three batches used for the calculation. In practice, it would be expected that many more than three stability studies would be available. The results of the example should not be used for deciding which method to use, only to understand how to take a set of data and calculate the OOT limits that might apply. The constant value K is chosen to be 3.0 for illustrative purposes. The OOT limits are determined using the equation:

$$\text{average change} \pm (K \times \text{standard deviation of the change})$$

Assume the data in Table I were obtained. A plot of these data is given in Figure 6. The columns of the data show how to calculate the points to be considered in the histogram. For example, for Batch 1, the change from previous method for the 12-month point is -3.1 ($95.3 - 98.4$); the change from previous (per unit of time) method for the 12 month point is -0.52 ($[(95.3 - 98.4) / 4] / 6$) if the units used is months; and the change from initial method for the 12 month point is -5.4 ($95.3 - 100.7$). The observed value method does not require any manipulation before evaluating the distribution. Thus the 12-month value is 95.3.

Batch	Time	Value	Change from previous method			Observed value
			Change from previous method	Change from previous method (per unit of time)	Change from initial method	
Batch 1	0	100.7				100.7
	3	98.4	-2.3	-0.77		98.4
	6	95.3	-3.1	-1.03		95.3
	12	95.3	-0.0	0.00	-5.4	95.3
Batch 2	0	100.7				100.7
	3	98.4	-2.3	-0.77		98.4
	6	95.3	-3.1	-1.03		95.3
	12	95.3	-0.0	0.00	-5.4	95.3
Batch 3	0	100.7				100.7
	3	98.4	-2.3	-0.77		98.4
	6	95.3	-3.1	-1.03		95.3
	12	95.3	-0.0	0.00	-5.4	95.3
Mean			-1.13	-0.38	-5.40	98.37
Standard deviation			1.50	0.50	1.50	1.50
Mean + 3 SD			1.23	0.50	1.50	100.10
Mean - 3 SD			-3.39	-1.38	-5.40	95.64

Table I: Detailed calculations for estimating analytical alert limits.

Note that the change from initial method must be evaluated at each time point. Because only three batches of data are included in the example, only three values are used to generate the limits for this method at each time point, which is very limited. Nonetheless, the calculations for the 12-month time point are provided only for illustrative purposes.