

Considerations When Determining a Routine Sample Size for a Retest Procedure

Jeffrey D. Hofer

This article discusses a straightforward approach that scientists may use to decide how many nonsuspect retest results are required to overcome an initial suspect result. The author compares the performance of various numbers of retests and proposes an approach that could be used to select the number of retests on the basis of appropriate risk levels. This methodology can provide a reasonable approach to selecting the correct number of retests to perform in investigations.

Jeffrey D. Hofer is a senior statistician in Global Statistical Sciences at Eli Lilly and Company, Lilly Corporate Center, Drop Code 3834, Indianapolis, IN 46285, tel. 317.276.9803, fax 317.277.5519, hofer_jeffrey_d@lilly.com.

During an evaluation of a pharmaceutical drug substance or drug product, a result may be obtained that is unusual when compared with the expected result. This kind of result is referred to as a *suspect result* and is defined as a result that does not meet a prespecified criterion. When a suspect result is observed, a thorough analytical and manufacturing investigation is conducted to assess whether the result is due to an assignable cause. In some cases, an assignable cause is found, and the appropriate actions are performed. In other cases, the investigations cannot find any assignable cause, and additional testing is required. Recently, regulatory agencies have focused intensely on how investigations of suspect data are performed and documented.

A recent FDA Draft Guidance for Industry, "Investigating Out of Specification (OOS) Test Results for Pharmaceutical Production," allows retesting to be performed to assess whether sufficient evidence exists to overcome the initial suspect result (1). One approach to overcoming suspect results is to generate sufficient data to demonstrate that the original suspect result was an outlier. This approach was discussed in detail in a previous study and will not be discussed in this article (2).

However, another valid but fundamentally different approach is to focus on the question of how many successive, nonsuspect retest results are required to overcome the initial suspect result. In this article, the term *overcome* is defined as the generation of sufficient data to determine that the material in question is acceptable relative to the prespecified criterion. The author addresses this alternative investigation approach and provides tools that can be used to determine the appropriate number of retests on the basis of the desired quality level.

Overview of approach and assumptions

As discussed previously, thorough analytical and manufacturing investigations that did not find any assignable cause must be conducted before developing a procedure for retesting. Then, scientists must decide the number of additional tests that must be performed to potentially overcome the initial suspect result.

Simply stated, the retests could be used in the following decision tree:

- If all K retest results are not suspect, then the initial suspect result is deemed to be overcome. This information can be forwarded to the quality assurance-quality control units for their

assessment of the appropriate disposition of the material.

- If one or more of the K retest results are suspect, then the retest criteria are not met.

In this article, the following factors are assumed when evaluating the properties of a retest procedure:

- The test sample from which the original and retest results are generated is homogeneous and representative of the material being evaluated.
- K retest results are independent observations.
- The definition of *suspect result* is the same for the initial result and the retest results. If not, the discussion summarizing the overall probability of meeting the prespecified criterion is not valid.

This article discusses the properties of the retest procedure, the overall probability of meeting the prespecified criterion, and the various quality standards that one could consider when determining how many retest results, K , should be routinely performed.

Properties of the retest

Assuming that K retest results are independent observations, the retest approach for any measurement method can be modeled by using a binomial distribution. The proportion of suspect results is denoted by p , in which $0 \leq p \leq 1$. Then, the proportion of nonsuspect results is denoted by q , in which $q = 1 - p$. When $100q\%$ nonsuspect results exist (i.e., $100p\%$ suspect results), the probability that K independent retests are performed with all the observed results being nonsuspect is q^K .

For example, assume that an initial suspect result is observed. Also assume that 25% of the results are truly suspect and that the retest strategy requires three nonsuspect results to overcome the initial suspect result. In this instance, the probability that the initial suspect result would be overcome is $(0.75)^3$ or 0.422. Thus, if 25% of the results are truly suspect and three nonsuspect retest results are required to overcome the initial suspect result, there is a 42.2% chance of overcoming an initial suspect result.

In another example, an initial suspect result is observed, and only 1% of the results are truly suspect. The retest strategy requires three nonsuspect results to overcome the initial suspect result. In this instance, the probability that the initial suspect result will be overcome is $(0.99)^3$ or 0.970. Thus, if 1% of the results are truly suspect and three nonsuspect retest results are required to overcome the initial suspect result, there is a 97% chance of overcoming an initial suspect result. Similar calcula-

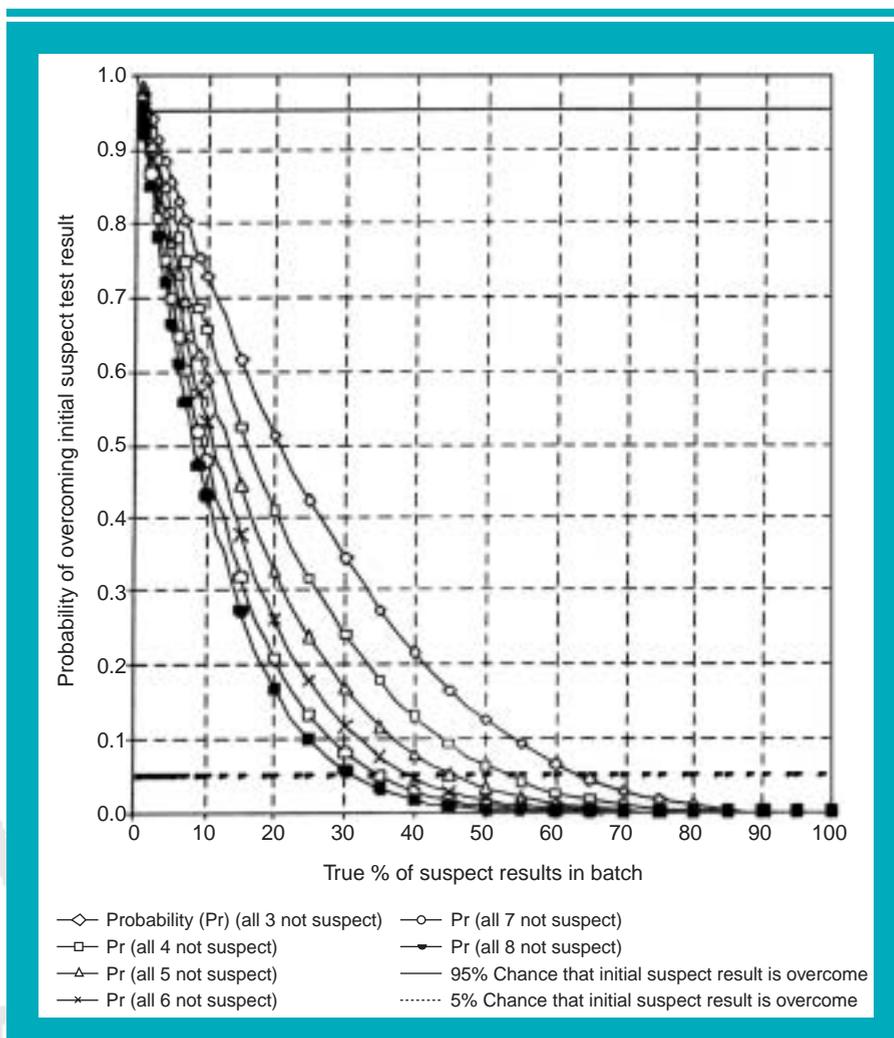


Figure 1: Probability of observing no suspect results in K retests versus true percentage of suspect results in the population.

Table I: Summary of true % of suspect results that have a specified chance of meeting retest criteria.

Number of retests, K	True % of suspect results with 95% chance of meeting retest criterion	True % of suspect results with 5% chance of meeting retest criterion
3	1.70	63.16
4	1.27	52.71
5	1.02	45.07
6	0.85	39.30
7	0.73	34.82
8	0.64	31.23

tions were performed for true percentages of suspect results ranging from 0 to 100% (see Figure 1). Note that the probabilities of overcoming the initial suspect result are unaffected if the initial suspect result is caused by an unidentified mistake and are only affected by the true proportion of suspect results in the remainder of the population.

As shown in Figure 1, for a fixed true percentage of suspect results, the probability of observing no suspect results in K

retests, or overcoming the initial suspect result, decreases as the number of additional tests is increased. Two areas of these curves can be compared to see the effect of the various numbers of retests.

The first area represents the true percentage of suspect results that would have a 95% chance of meeting the retest criterion when K retests are performed. This can be thought of as the producer's risk because there are very few suspect results but still a 5% chance of not meeting the retest criteria.

The second area represents the true percentage of suspect results that would have only a 5% chance of meeting the retest criterion when K retests are performed. This can be thought of as the consumer's risk because there is a high percentage of suspect results but still a 5% chance of meeting the retest criterion.

These quantities can be calculated using the following equations, in which K represents the number of retests to be performed (see Table I):

$$\text{True \% of suspect results with 95\% chance of meeting retest criteria} = 1 - (0.95)^{1/K}$$

$$\text{True \% of suspect results with 5\% chance of meeting retest criteria} = 1 - (0.05)^{1/K}$$

For example, if five retests are performed and if there are truly 1.02% suspect results, a 95% chance exists of overcoming an initial suspect result (if all five retest results are nonsuspect). However, if there are 45.07% suspect results, only a 5% chance exists of overcoming an initial suspect result.

Table I shows that the more retests that are performed, the lower the percentage of suspect results needs to be to achieve the same frequency of overcoming an initial suspect result. For example, if only three nonsuspect retests are required, there is a 95% chance of overcoming an initial suspect result with 1.7% suspect results. However, if eight nonsuspect retests are required, there can only be 0.64% suspect results to have a 95% chance of overcoming the suspect result.

The overall probability of meeting the prespecified criterion

The probability of overcoming an initial observed suspect result is only part of the overall probability of meeting the prespecified criterion. The other part requires considering how often an initial suspect result might be observed and how often

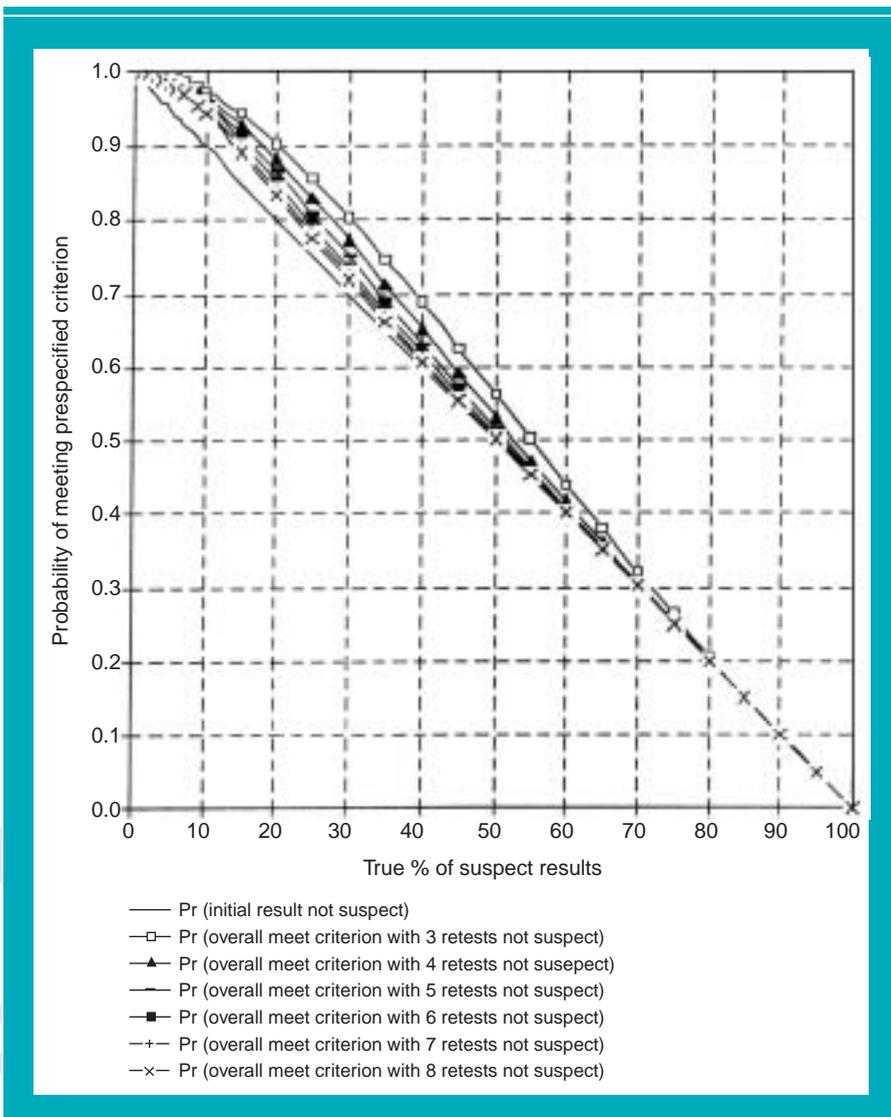


Figure 2: Initial and overall probability of meeting the prespecified criterion using the procedure with K retests versus the true percentage of suspect results.

an investigation would be initiated. For a given situation (assuming no analytical mistakes exist), if there are truly $100p\%$ suspect results, then $100(1 - p)\%$ of the time the initial result is not suspect and the result meets the prespecified criterion. Thus, $100p\%$ of the time a suspect result is observed. Figure 1 shows the probability of overcoming the suspect result in the retest investigation as a function of p and K .

So what is the overall probability of meeting the prespecified criterion when there are truly $p\%$ suspect results? To summarize, a $100(1 - p)\%$ chance exists that the initial result meets its prespecified criterion, and of the $p\%$ of the time that the retest investigation occurs, the percentage of the time that the initial result is overcome is $100(1 - p)^{K\%}$.

Thus, the overall probability of meeting the prespecified criterion is

$$100(1 - p)\% + 100p(1 - p)^{K\%}$$

For example, when 25% suspect results exist in which $K = 3$

retests are performed, there is a 75% chance that the initial result meets its prespecified criterion. Of the 25% of the time that the retest occurs, the initial result is overcome 42.2% of the time. Thus, when three retests are performed, the overall probability of meeting the prespecified criterion with 25% suspect retests is

$$75\% + (25\%)(42.2\%) = 85.6\%$$

In another example, when 1% suspect results exist in which $K = 3$ retests are performed in the retest investigation, there is a 99% chance that the initial result meets its prespecified criterion. Of the 1% of the time that the retest investigation occurs, the initial result is overcome 97% of the time. Thus, when three retests are performed, the overall probability of meeting the prespecified criterion with 1% suspect results is

$$99\% + (1\%)(97\%) = 99.97\%$$

Similar calculations were performed for true percentages of suspect results ranging from 0 to 100%. Figure 2 shows both the probability of initially meeting the prespecified criterion and the overall probability of meeting the prespecified criterion when using the two-stage approach with various numbers of retests.

Figure 2 shows that it is difficult to clearly identify the effect that the various numbers of retests have on the overall probability of meeting the prespecified criterion because of the frequency with which the initial result would be nonsuspect and no retesting would be initiated. Figure 3 shows the increase in the probability of meeting the prespecified criterion as a result of retesting beyond the probability of meeting the prespecified criterion solely on the basis of the initial result.

Note that none of the retest approaches increases the probability of meeting the prespecified criterion by more than 11% beyond the probability of meeting the prespecified criterion solely on the basis of the initial result. This maximum increase occurred when ~25% suspect results existed and $K = 3$ retests were performed.

Selecting K depends on choice of quality standard

Choosing the routine number of retests to perform, K , depends on the desired quality standard. A myriad of options are possible with the information that has been presented in this article, but only a few will be discussed in the following text.

Controlling the probability of meeting the retest criterion to $\leq 5\%$

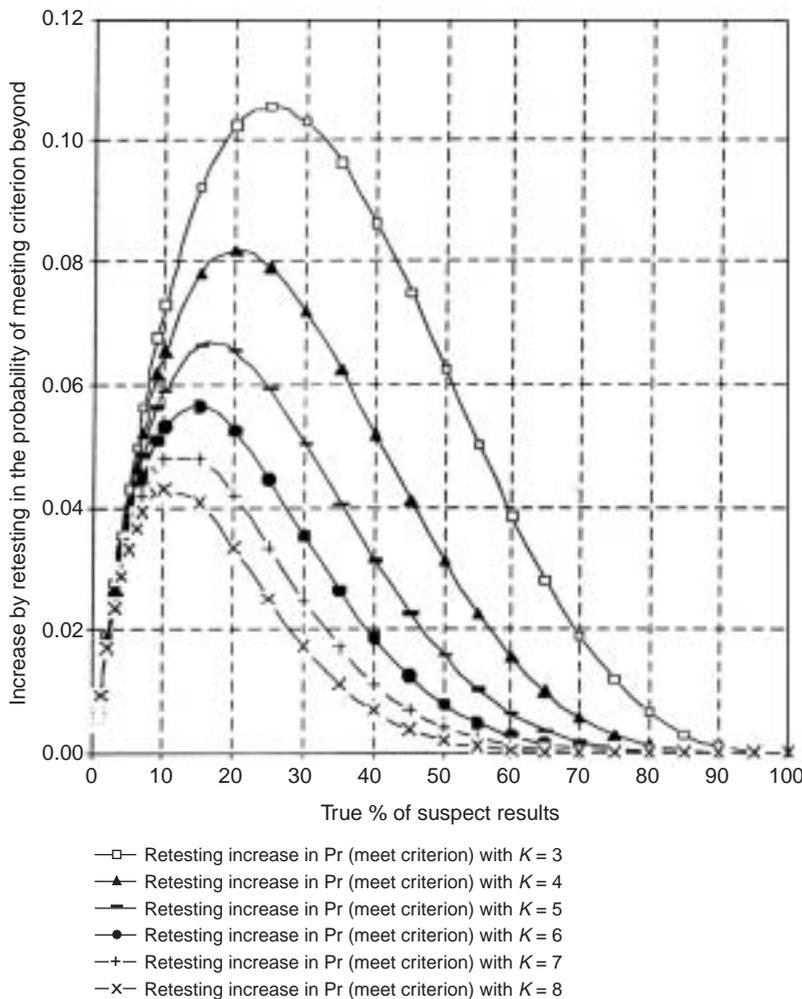


Figure 3: The increase in probability of meeting the prespecified criterion as a result of performing K retests versus the true percentage of suspect results.

of the time when the true mean is equal to either the lower or upper limit of the prespecified criteria. In this situation, a distribution will have 50% of its results deemed suspect. One quality standard includes having a retest procedure that, when there are 50% suspect results, no more than a 5% chance exists of having the criteria met in the retest investigation.

Because half the results are suspect and half are not, this situation could be thought of as calculating the probability of how many times a person can flip a coin that will turn up heads K successive times. That is, one would choose the first K in such a way that 0.5^K was less than 0.05, which would occur with $K = 5$ retests. The probability of observing five consecutive non-suspect results when there are 50% suspect results is 0.031. Thus, choosing this quality standard would lead to the selection of five retests for routine performance.

This does not imply that the overall probability of meeting the prespecified criterion is 3.1%, but rather that only a 3.1% chance exists of meeting the retest criterion. The overall probability of meeting the prespecified criterion is 51.6% using the calculation approach described previously.

RECUSTOM EPRINTS

Reprints of *Pharmaceutical Technology* articles, advertisements, news items or special announcements are available through Advanstar Marketing Services. Customized to meet your specific needs, reprints are highly effective when you use them to:



- ♦ Develop direct-mail campaigns
- ♦ Provide product/service literature
- ♦ Create trade show distribution materials
- ♦ Present information at conferences and seminars

ARTICLES
NEWS ITEMS
ADVERTISEMENTS

Extend your coverage to your website.
Custom reprint packages include an E-Print of the same article to post on your website.

Pharmaceutical Technology®

MARY CLARK
ADVANSTAR MARKETING SERVICES
800-822-6678 ext. 226
541-984-5226 Fax: 541-686-5731
Email: mclark@advanstar.com

Selecting the number of retests that provides the retest investigation with the desired discrimination properties. As previously mentioned, calculations of the probability of overcoming an initial suspect result were performed for various numbers of retests. After a review of the values in Table I, key scientists should select the number of retests that allows them to discriminate between poor-quality and high-quality material. That is, the scientists can determine how low the true percentage of suspect results should be while maintaining a high probability of meeting the prespecified criterion K successive times during the retest investigation. Simultaneously, scientists can evaluate how high the true percentage of suspect results can be before there is a very low chance of meeting the prespecified criterion K successive times.

For example, suppose key scientists want a 95% chance of meeting the retest criterion when there are 1% suspect results, and simultaneously that there would be less than a 5% chance of meeting the retest criterion with 50% or greater suspect results. With these constraints, the choice of five retests would meet this requirement because when there are 1.02% suspect results, the retest criterion will be met 95% of the time, and if there are 45.07% suspect results, there will be only a 5% chance of meeting the retest criterion.

In contrast, if three retests were selected, and 1.7% suspect results exist, the retest criterion would be met 95% of the time. If as many as 63.16% suspect results exist, the retest criterion would be met 5% of the time.

Furthermore, if seven retests were selected, and 0.73% suspect results exist, the retest criterion would be met 95% of the time. If as many as 34.82% suspect results exist, the retest criterion would be met 5% of the time.

Conclusion

This article provides a rationale to assist in determining the number of nonsuspect retest results that must be generated to overcome an initial suspect result. Scientists should choose the number of retests that will allow them to adequately discriminate between material of acceptable and unacceptable quality.

Acknowledgments

The author would like to recognize and thank Gregory Davis, Jeffrey Gelwicks, Pedro Jimenez, Kristi Griffiths, Doris Weisman, and the manufacturing and development (MADSIG) statisticians at Eli Lilly and Company for their input during the development and compilation of this article.

References

1. FDA, "Draft Guidance for Industry: Investigating Out of Specification (OOS) Test Results for Pharmaceutical Production," Center for Drug Evaluation and Research, September 1998.
2. J.D. Hofer and E.C. Rickard, "A Guide to Determining Sample Sizes and Outlier Criteria in Analytical Investigations," *Pharm. Technol.* **25** (3), 52-66 (2001). **PT**

For Client Review Only. All Rights Reserved. Advanstar Communications Inc. 2003