

# Examination of Selection of Immediate Release Dissolution Acceptance Criteria

Jeffrey D. Hofer,<sup>1</sup> and Vivian A. Gray<sup>2</sup>

<sup>1</sup> Lilly Research Laboratories, Indianapolis, IN

<sup>2</sup> President, V.A. Gray Consulting, Inc., Hockessin, DE

email correspondence to: hofer\_jeffrey\_d@lilly.com or  
vagrays@rcn.com, www.vagraysconsulting.com

This article has been reprinted with permission granted from United States Pharmacopeial Convention, Rockville, MD. Reference *Pharmacopeial Forum* 29(1) pages 335-340 for original publication.

## Abstract

Controversy and approval delays often arise between a sponsor company and regulatory agencies over the establishment of dissolution acceptance criteria. For a US submission, the dissolution specification that is established is based on the acceptance criteria in USP Dissolution General Chapter <711>. When defining the acceptance criteria, the sponsor must propose a 'Q' value and a time point (e.g., 30 minutes or 45 minutes) at which point the data will be evaluated against the appropriate criteria. In many instances, there is disagreement between the sponsor and the FDA on the appropriate values for the collection time point and the value of 'Q'.

This paper will illustrate the very conservative nature of the Stage 1 acceptance criteria relative to that of the other stages. This paper demonstrates that a major issue that must be addressed when approaching this topic is the large difference in the discriminatory ability of the initial and subsequent stages of the dissolution test. Another issue brought forth is the traditional establishment of the Q value only in 5 unit increments. This paper will illustrate via theory and examples the problems presented by the above two issues. Finally, this article will present an alternative data-driven approach that could be used to arrive at potential dissolution specifications.

## Introduction

Prior to discussing this issue, it is important that one be familiar with the immediate release acceptance criteria of the USP Dissolution test, shown below in *Table 1*.

For a US submission, in order to define the dissolution specification, the sponsor must propose a 'Q' value and a time point (e.g., 30 minutes or 45 minutes) at which point the data will be evaluated against the appropriate criteria. In many instances, there is disagreement between the sponsor and the FDA on the appropriate values for the collection time point and the value of 'Q', leading to approval delays and controversy between the sponsor and regulatory agency.

This paper will illustrate the very conservative nature of the Stage 1 acceptance criteria relative to those of the other stages. This paper demonstrates that a major issue that must be addressed when approaching this topic is the large difference in the discriminatory ability of the initial and subsequent stages of the dissolution test. Another issue brought forth is the traditional establishment of the Q value only in 5 unit increments. This paper will illustrate via theory and examples the problems presented by the above two issues. Finally, this article will present an alternative data-driven approach that could be used to arrive at potential dissolution specifications.

## Discussion

A review of *Table 1* demonstrates that the first stage of the USP dissolution test consists of testing six dosage units. If all of the dosage

units are greater than or equal to Q+5, then the dissolution test criteria are met and the test is passed.

However, if this criterion is not met, six additional dosage units are tested and compared to the acceptance criteria for the twelve dosage units. To pass at the second stage, the average of the twelve dosage units must be equal to or greater than Q and no dosage unit can be less than Q-15%.

If both of the above criteria are not met at the second stage, the final stage of testing is performed. Twelve additional dosage units are evaluated, providing a total of twenty-four results. To pass at this final stage of testing, the average of the twenty-four dosage units must be equal to or greater than Q, not more than two dosage units can be less than Q-15%, and no dosage unit can be less than Q-25%.

**Table 1. USP Dissolution Acceptance Criteria**

Stage	Number of Dosage Units Tested	Pass if
1	6	No dosage unit is less than Q+5%
2	6	Average of the twelve dosage units $\geq$ Q% And No dosage unit is less than Q-15%
3	12	Average of the twenty-four dosage units $\geq$ Q% And Not more than two dosage units are less than Q-15% And No dosage unit is less than Q-25%

As we will demonstrate, there are large differences in the test discrimination between the first stage and the subsequent stages. As seen above, for a given Q, to pass the dissolution test at Stage 1, each of the six results must be greater than or equal to Q+5%. To estimate how often the test would pass at Stage 1, it is necessary to estimate the probability that an individual dosage unit is greater than or equal to Q+5%. This probability can be estimated from the data collected. Refer to this probability of an individual dosage unit being greater than or equal to Q+5% as p. The probability of passing the dissolution test at Stage 1 is the probability that all six results are greater than or equal to Q+5%, which can be calculated as  $p^6$ . Thus, the probability that Stage 2 is required is  $1 - p^6$ .

For example, assume that 90% of the individual results are estimated to be greater than or equal to Q+5% (i.e.,  $p=0.90$ ). Then, the probability of passing the dissolution test at Stage 1 can then be calculated as  $0.9^6$ , or 0.53. So, there is a 53% chance of passing the dissolution test at Stage 1 thus, the probability that Stage 2 is required is  $1 - p^6$ , or 47%.

Figure 1 below shows the probability that Stage 2 testing will be required as a function of the percentage of the distribution of results that are greater than Q+5%. As illustrated, unless there is a high percentage of the individual results greater than Q+5%, Stage 2 testing will frequently be required. For a distribution with 85% of the individual results greater than Q+5%, Stage 2 testing would be required 65% of the time. In fact, for a sample with 70% of individual results greater than Q+5%, Stage 2 testing will be required 90% of the time.

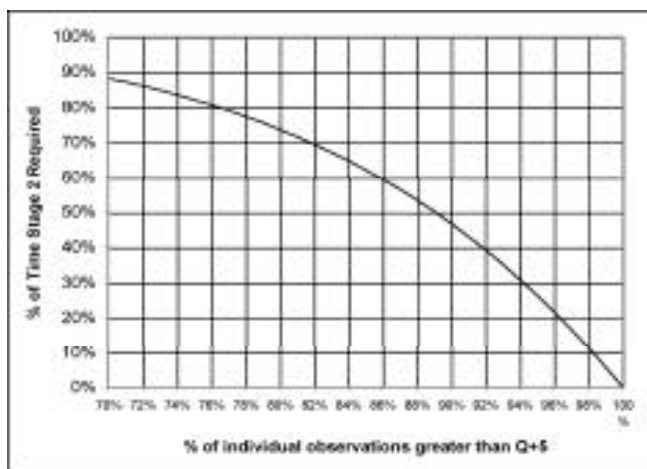


Figure 1. Plot of Stage 2 Frequency as a function of percent of individual results greater than Q+5%

If Stage 2 testing is required, Table 1 shows that passing at Stage 2 depends upon whether the observed average of the twelve results is greater than or equal to Q and whether any

of the results are less than Q-15%. If the criteria are not met at Stage 2, twelve additional dosage units are tested and the test is passed at Stage 3 if the observed average of the twenty-four results is greater than or equal to Q and not more than two individual results are less than Q-15% and no result is less than Q-25%.

A simulation study was performed to evaluate the percentage of time that the dissolution test criteria would be met for samples of varying quality. The simulation assumed that the data followed a normal distribution. This is felt to be a reasonable assumption for the typical collection time points where immediate release dissolution specifications are routinely established (usually 30, 45, or 60 minutes). The true mean of the sample from which the test article was drawn was assumed to vary from 6% below Q to 8% above Q. The standard deviation of the sample results was assumed to be 2, 4, or 6. Larger standard deviations were evaluated but it was felt that the maximum true standard deviation of interest for a collection time point where there was a dissolution specification was 6.

Figure 2 below shows a plot of the percentage of tests that would pass either at Stage 1, at Stage 2, or at Stage 3 as a function of the percentage of the individual results that are greater than Q. It was shown earlier that the probability of passing the dissolution criteria at Stage 1 is a function of the percentage of results greater than Q+5%. However, the ability to pass the dissolution test at the later stages is primarily a function of the percentage of results greater than Q%. Thus, for comparability among the stages, Figure 2 uses the percent of individual results greater than Q% as the x-axis to illustrate the large disparity between the test requirements at Stage 1 and thereafter. A review of the figure shows that for samples where there is a very high chance of passing at Stage 2 there is very little chance of passing the test at

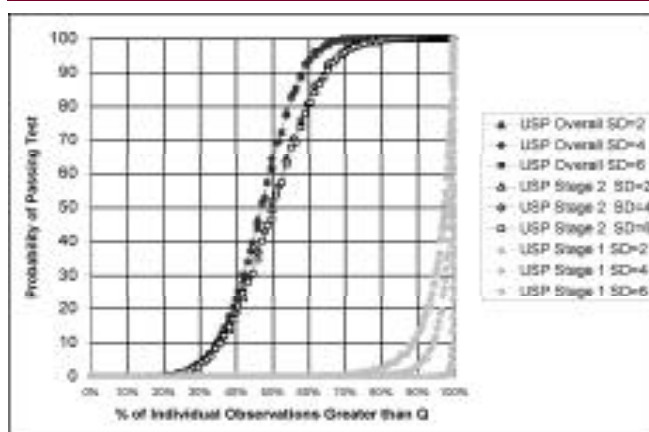


Figure 2. Plot of Probability of Passing the Dissolution Test at Stage 1, Stage 2 or Stage 3 as a function of percent of individual results greater than Q%

## Dissolution Acceptance Criteria... continued

Stage 1. This is because the Stage 1 requirements are that nearly all of the individual results fall above  $Q+5\%$  while the Stage 2 requirements primarily depend upon whether the average is slightly greater than  $Q\%$ .

Figure 2 also demonstrates that the percentage of passing tests at Stage 2 or 3 is independent of the standard deviation for those standard deviations examined in the simulation. This is because with true standard deviations of 6 or less, the criteria on individual results is very unlikely to come into play in the passing or failing of the test and thus the test simplifies down to a comparison of the observed average to  $Q$ .

A brief discussion of why the above observation is true is provided. Reviewing Table 1, it is noted that one way to fail the dissolution test is to observe three or more results less than  $Q-15\%$  out of the twenty-four results. How often would this be expected to occur while the average of the results is greater than  $Q$ ? Consider the situation where the true distribution has a mean of  $Q$  and a standard deviation of 6. For an individual result to be less than  $Q-15\%$ , it needs to be 2.5 or more standard deviations below the mean, which should occur only 0.62% of the time. Thus, in such a situation, one should observe three or more results less than  $Q-15\%$  out of a sample of twenty-four only 0.04% of the time. Thus, it is highly unlikely that one would fail the dissolution test due to this criterion. One is much more likely to fail due to the observed average being less than  $Q$ , which should occur 50% of the time by chance at each Stage if the true mean is  $Q$ . In addition, one result of the twenty-four less than  $Q-25\%$  is very unlikely to occur while the true mean is at, or above,  $Q$  since this would be an observation that, in this example, is more than four standard deviations below the mean.

Similar calculations can be performed for a true mean of  $Q$  and a true standard deviation of 8. Consider the situation where the true distribution has a mean of  $Q$  and a standard deviation of 8. For an individual result to be less than  $Q-15\%$ , it needs to be 1.875 or more standard deviations below the mean, which should occur only 3.04% of the time. Thus, in such a situation, one should observe three or more results less than  $Q-15\%$  out of a sample of twenty-four only 3.53% of

the time. Thus, it is highly unlikely that one would fail the dissolution test due to this criterion. One is much more likely to fail due to the observed average being less than  $Q$ , which should occur 50% of the time at each Stage if the true mean is  $Q$ . In addition, one result of the twenty-four less than  $Q-25\%$  is very unlikely to occur while the true mean is at, or above,  $Q$  since this would be an observation that, in this example, is more than three standard deviations below the mean.

Figure 2 also shows that as long as 80% of the distribution of individual results is greater than  $Q$ , the test should be met at Stage 2 while as long as 70% of the distribution is greater than  $Q$  the test will virtually always be met by Stage 3. If the true mean is  $Q$ , there is a 50% chance of passing at Stage 2 while there is a 62% chance of passing after Stage 3. For samples where only 33% of the individual results are greater than  $Q$ , there is only a 5% chance of passing the test.

As can be seen from Figures 1 and 2, there is a tremendous degree of disparity between the test discrimination at Stage 1 and thereafter. In order to routinely pass the USP dissolution test at Stage 1, virtually all of the individual results from the true distribution need to be greater than or equal to  $Q+5\%$ . This implies that the mean of the data from which the individual samples are drawn needs to be at least three standard deviations above  $Q+5\%$ . However, if the Stage 1 criteria are not met, only about 85% of the individual results need to be greater than  $Q$  to routinely pass the USP dissolution test at Stage 2. This implies that the mean of the data from which the individual samples are drawn needs to be at least one standard deviation above  $Q\%$  in order to routinely pass after Stage 2.

To summarize the above information, a summary table may prove helpful. Tables 2 and 3 show the following for  $Q$  values of 75 and 85, respectively:

- the minimum mean necessary for essentially no Stage 2 testing for different standard deviations
- the minimum mean that will lead to Stage 2 testing being required 90% of the time
- the minimum mean necessary for essentially 100% passing after Stage 2.

**Table 2.  $Q=75$  Summary of Means leading to Stage 2 testing percentages and passing after Stage 2**

True Standard Deviation	$Q=75$ Minimum Mean for No Stage 2 Testing	$Q=75$ Minimum Mean that will lead to Stage 2 testing approximately 90% of the time	$Q=75$ Minimum Mean for passing after Stage 2 essentially 100% of the time
2%	86%	81.0%	77%
3%	89%	81.5%	78%
4%	92%	82.0%	79%
5%	95%	82.5%	80%
6%	98%	83.0%	81%

The accuracy of the results summarized in Tables 2 and 3 (particularly the second column showing the minimum mean required for essentially no Stage 2 testing) are dependent upon the validity of the assumption that the data are normally distributed.

Thus, as can be seen from Table 2, for a fixed standard deviation there is a very large discrepancy between the

**Table 3. Q=80 Summary of Means leading to Stage 2 testing percentages and passing after Stage 2**

True Standard Deviation	Q=80 Minimum Mean for No Stage 2 Testing	Q=80 Minimum Mean that will lead to Stage 2 testing approximately 90% of the time	Q=80 Minimum Mean for passing after Stage 2 essentially 100% of the time
2%	91%	86.0%	82%
3%	94%	86.5%	83%
4%	97%	87.0%	84%
5%	100%	87.5%	85%
6%	103%	88.0%	86%

requirements on the true mean in order to routinely pass at Stage 1 and at Stage 2. This large difference is often the source of controversy over the establishment of the 'Q' value since samples with true means just slightly above Q will pass after the Stage 2 testing while samples need to have means much greater than Q+5% to pass after only Stage 1 testing.

For the scenarios examined in the simulations, Stage 3 is seen to only slightly increase the ability to pass samples drawn from distributions where less than 50% of the individual results were less than Q. The primary increase in pass percentage between Stages 2 and 3 was observed for samples coming from distributions where between 50% and 84% of the individual results were greater than Q (or for distributions where the mean was between Q and Q+1s). It is also observed that distributions with true means one or more standard deviations below Q have less than a 5% chance of passing the dissolution test.

**Example 1 – What are the True Mean requirements to easily pass at Stage 1 or 2?**

A specific example is presented to put the above general concepts into specific terms. Assume that an acceptance criteria of Q=80% at 30 minutes has been established for a compound. To routinely pass the dissolution test at Stage 1, the mean of the data from which the individual samples are drawn needs to be at least three standard deviations above Q+5%. Thus, if the standard deviation is 4%, then the mean needs to be at least 97% {i.e.,  $Q+5+(3*SD)$ } in order to nearly always pass at Stage 1. If the mean is 87.0% or less, there will be Stage 2 testing approximately 90% of the time. Additionally, to routinely pass the test after Stage 2, the mean of the data from which the individual samples are drawn needs to be at least 84%.

**Example 2 – Establishing an 'Acceptable' Q value and Collection Time Point**

Consider the following example that illustrates the difficulties in establishing a Q value and collection time point specification that both the regulatory body and the sponsor deem acceptable. Assume that a tablet product has been developed and is going to be submitted to the regulatory agency.

Assume that dissolution profile data have been collected for clinical and registration stability batches. Assume further that the profiles are very similar and that there is no change in the dissolution profile during the long-term stability of the product. Thus, all of the data can be combined to obtain the most accurate estimates of the average and standard deviation at the different collection time points. Usually, there are only one or two candidate time points for consideration and one will propose a Q value of 75 or 80 at one of these times as the proposed dissolution specification. Assume that the following summary statistics are obtained for this example as shown in Table 4.

Note that the calculations for the percent of individual results less than Q+5% have been changed to the percent of individual results less than Q+4.5%. This was done to account for the rounding aspect of the dissolution test when comparing individual results to their acceptance criteria. For example, if the value of Q is 80, then an individual result of 84.5% would meet the acceptance criteria of being greater than or equal to Q+5%.

As can be seen from Table 4, there is quite a challenge in establishing a Q value and collection time point specification that is acceptable to both the regulatory agency and the sponsor due to the wide disparity in the required amount of predicted Stage 2 testing. Reviewing Table 4, the sponsor will likely propose a specification of Q=80% at 45 minutes. However, it is noted that if future lots have dissolution profile

**Table 4. Summary Statistics for Hypothetical Example**

Collection Time Point, Minutes	Number of Individual Results	Average	Standard Deviation	Potential Q-value	% of individual results predicted to be less than Q+4.5%	Predicted Frequency of Stage 2 Testing
30	252	90.20	4.36	75	0.74%	4.36%
30	252	90.20	4.36	80	9.61%	45.48%
45	252	96.78	3.51	75	0.00%	0.00%
45	252	96.78	3.51	80	0.03%	0.17%



# Dissolution Acceptance Criteria... continued

characteristics similar to the lots analyzed to obtain the summary statistics there will be a very low occurrence of Stage 2 testing. It is observed that a specification of  $Q=75\%$  at 30 minutes could also be proposed and lead to a relatively low frequency of Stage 2 testing while a specification of  $Q=80\%$  at 30 minutes would lead to quite extensive Stage 2 testing. It is not clear what the reaction of regulatory agencies would be to the proposed specification given the very low expected frequency of Stage 2 testing.

To summarize, two challenges that exist in establishing dissolution specifications are highlighted from this example. The first is that data is only collected at a few collection time points and thus, the specifications will need to be established at one of these time points. The second is that the  $Q$  values typically established are traditionally only set in 5-unit increments. Even if there were an agreed upon level of Stage 2 testing between industry and regulatory agencies, the above two constraints make it nearly impossible to meet this requirement except in rare chance occurrences.

For example, assume that it could be agreed by all that 20% Stage 2 testing would be desired. This would require that no more than 4% of the individual distribution results are less than  $Q+5\%$  (or  $Q+4.5\%$  taking into account the normal rounding procedures). This Stage 2 testing frequency would be achieved if the true mean of the distribution were 1.75 above  $Q+5\%$  (or  $Q+4.5\%$  as mentioned above). Thus, the  $Q$ -value to achieve this Stage 2 testing could be obtained by subtracting the quantity '5+1.75' (or the quantity '4.5+1.75') from the overall average. Table 5 summarizes these 'Stage 2 Testing @ 20%'  $Q$  values for the above example.

If the restriction is established that the  $Q$  value will not exceed 80% as is usually done, then the specification that could be perhaps agreed upon is  $Q=78\%$  at 30 minutes. Alternatively, the specification could be established at  $Q=80\%$  at 45 minutes with the acknowledgement that little, if any, Stage 2 testing will be required unless there is a change in the dissolution profile for future batches from the data originally submitted.

It is noted that maintenance of Stage 2 testing at the initial predicted frequency will be difficult as small changes in the mean or variability can have a dramatic impact on the Stage 2 testing percentage. This sensitivity

was displayed in detail in Figure 1.

As an example of this sensitivity, consider a situation where a  $Q=80$  had been established. Assume that originally the true standard deviation was 4 and that the true process average originally was 91.0. Thus, at the time of submission and specification establishment, one would predict that Stage 2 testing would be required approximately 25% of the time since the observed mean is 1.645 standard deviations above  $Q+5\%$ .

However, if the mean would fall to 89.3 with the standard deviation remaining at 4, then Stage 2 testing would be required approximately 50% of the time since the observed mean would then only be 1.2 standard deviations above  $Q+5\%$ . An increased requirement for Stage 2 testing of 50% could also occur if the mean were to remain at 91.0 but if the standard deviation were to increase to 5.4.

In this example, it is possible that the Stage 2 testing frequency could also decrease to virtually 0%. This could occur if the mean were to be increased to 96.5 while the standard deviation remained at 4 or if the mean were to remain at 91.0 while the standard deviation decreased to 2.2. In both of these situations, the mean would be three standard deviations above  $Q+5\%$ .

To summarize, the general dissolution specification methodology outlined in this paper is easily adaptable for whatever frequency of Stage 2 testing is agreed upon by the interested parties. It is the strong opinion of the authors that without such an agreement, the establishment of dissolution specifications will continue to be fraught with controversy.

## Conclusions

A great deal of controversy often arises between a sponsor company and regulatory agencies over the establishment of dissolution acceptance criteria. Much of this controversy revolves around the establishment of the  $Q$  value and the collection time point. A major issue that must be addressed when approaching this topic is the large difference in the discriminatory ability of the initial and subsequent stages of the dissolution test. Another issue is the traditional establishment of the  $Q$  value only in 5 unit increments. This article has illustrated via theory and examples the problems presented by the above two issues. This

article has presented an alternative data-driven approach that could be used to arrive at potential dissolution specifications. This approach is predicated on the assumption that an agreement can be reached between the industry and regulatory agencies on the acceptable frequency of Stage 2 testing.

**Table 5. Q Values to yield 20% Stage 2 Testing for Hypothetical Example**

Collection Time Point, Minutes	Number of Individual Results	Average	Standard Deviation	Q-value to achieve Stage 2 Testing Frequency of 20%	Q-value in whole units to achieve Stage 2 Testing Frequency of 20%
30	252	90.20	4.36	78.07	78
45	252	96.78	3.51	86.14	86