A Recommendation for Reduced Dissolution Calibration Testing[a]

PhRMA Dissolution Test Working Group;
The PhRMA Subcommittee on Dissolution Calibration
The members of the PhRMA Subcommittee on
Dissolution Calibration are: Onyenoba Chukwumerije,
Stephen Brune, Kathleen Hippeli, Al Kentrup, Lisa Martin,
Mary Oates, Dennis Stelzer, and David Whiteman.

Introduction

In May 96, a subcommittee of the PhRMA Dissolution Committee was formed to evaluate alternate dissolution apparatus suitability test requirements. The evaluation of alternate test requirements was stimulated by knowledge that lot L is the final lot of 50 mg Prednisone Calibrator Tablets to be manufactured by the current manufacturer and growing concerns throughout industry and regulatory agencies over the appropriateness of the current test requirements (1-6).

Discussions on the appropriateness of the current test requirements have focused on the usefulness of all eight suitability tests in providing unique and essential information (1-3); the apparent tablet to tablet variabilities found within the lots of calibrator tablets; the availability of improved dissolution apparatus and reliable measurement devices (4); and the potential adoption of alternate calibrator tablets (e.g., NCDA#2 prednisone tablets) (4-6). While the subcommittee is seriously considering several alternatives arising from the factors listed above, the USP and the subcommittee desired an immediate change to the testing requirements. Reducing testing requirements will extend the life of the current calibrator tablets and lengthen the time available for the identification and implementation of alternate suitability testing requirements. Thus, the most appropriate action at this time was to evaluate the ability to reduce the amount of suitability testing using the same 50 mg prednisone and 300 mg salicylic acid tablets currently in use.

This idea is not new. In fact, it was recently suggested in a Pharmacopeial Forum stimuli article by Richard Lindauer (1). Mr. Lindauer suggested that based on the results of the 1994 collaborative study, the required eight suitability tests could have been reduced to four tests "without losing essential information on the suitability of a laboratory's dissolution testers." In this article, only the 1994 collaborative data was reviewed, thus the proposed reduced testing scheme was specific to the 1994 calibrator tablets. In addition, the four selected tests were chosen based on their confirmed sensitivity to deaeration of the media.

This report describes a different statistical evaluation aimed at identifying a reduced dissolution suitability testing scheme and reports the results of that evaluation. Specifically, data from the collaborative studies conducted in 1989, 1991, and 1994 were used in the analysis.

Data Analysis

The basic premise to our approach was that only those suitability tests that lend unique and essential information should be required. Those tests that consistently yield passing results are most likely not discriminating enough to be of value. Conversely, we must assume that tests by which many laboratories yield out-of-calibration or failing results are more highly discriminating and potentially indicative of a problem (e.g., insufficient deaeration, vibration, paddle or vessel mis-alignment, etc.). Furthermore, any reduced testing scheme needs to detect a majority of these failing or out-of-calibration baths. The question of what constitutes a majority was answered during the data review process and is described later in this paper.

The best source of information on the frequency of failures for the suitability tests were the PMA (now PhRMA) reports on the collaborative studies of dissolution calibrator tablets (9-10). Since we were interested in evaluating data across different lots of calibrator tablets without regard to cited deaeration sensitivities (1), all collaborative data from 1989, 1991, and 1994 were examined. While it is true that deaeration was increasingly controlled from 1989 to 1994, all data were handled with equal weight. Furthermore, all baths with out-of-calibration or failing results were identified based on the statistically-determined acceptable ranges for percent tablet dissolved

5

A Recommendation for Reduced Testing...cont.

Table 1. Assigned Test Numbers for Dissolution Suitability Tests

Test Number	Calibrator	Apparatus	Speed
1	Prednisone	Basket	50 rpm
2	Prednisone	Basket	100 rpm
3	Prednisone	Paddle	50 rpm
4	Prednisone	Paddle	100 rpm
5	Salicylic Acid	Basket	50 rpm
6	Salicylic Acid	Basket	100 rpm
7	Salicylic Acid	Paddle	50 rpm
8	Salicylic Acid	Paddle	100 rpm

which are not necessarily the same as the final USP-approved acceptable ranges (11).

To facilitate data analysis, the eight suitability tests were numbered 1 through 8 as shown in Table 1. For each collaborative study, failing results were identified by laboratory and test

Table 2. Laboratories Reporting Out-of-Calibration Dissolution Results for Three Collaborative Studies

Test Number	OUT-OF-CALIBRATION LABORATORIES		
	1989[1]	1991[2]	1994[3]
1	G	5, 22	2, 17, 21
2	O	1, 2, 4, 5, 6, 18	2, 6, 7, 11, 12, 25, 27
3	H, J, L, N	6, 9, 15, 16	9, 14, 22, 25
4	H, L, N	5, 6	9
5	O	1, 2, 19	12
6	H, O	11	11, 12, 16, 20, 27
7	C, H, J, L	2, 4, 5, 9	2, 16, 20, 23
8	E	4, 5, 22	2, 4, 9, 18, 19, 27

1. Eighteen laboratories participated (A- R)
2. Twenty-two laboratories participated (1 - 22)
3. Twenty-seven laboratories participated (1 - 27)

number (see Table 2). The information contained in Table 2 was used to identify which single test detected the most of the suitability failures and to identify which combination(s) of two, three, and four tests identified the most suitability failures.

In addition to determining the overall best combinations of tests, a second approach was also evaluated. Although eight suitability tests are defined, dissolution baths are often only calibrated for their intended use (8). Baths intended for use with baskets or paddles only require four suitability tests. The ideal reduced testing scheme would preserve the ability to test baths for only their intended use and simply combine those apparatus-specific tests for full testing. Reduced testing for either baskets or paddles may involve three, two, or one test.

Tables 3 and 4 (page 7) contain information on the percentage of failures detected using combinations of two tests for baskets and paddles, respectively. Note that in each case the total number of failures includes only failures found using baskets (Table 3) or paddles (Table 4). (Assumes that, in this case, only basket or paddle failures are significant to basket or paddle use, respectively.)

Results and Discussion

Results from the first approach in which all combinations were evaluated, without regard to the type of apparatus, are given in Table 5 (page 8). The best combinations of one, two, three and four tests are identified by test numbers (refer to Table 1) and the percentages of failures detected are reported. Note that a minimum of three suitability tests was required to detect at least 75% of all bath failures. Using a combination of four suitability tests, 89% of all failures were detected. Further increases to the number of tests in the combination were not evaluated since the "cost" (including time and labor) of each additional test was not justified by the information gained (i.e., combinations of the best five to seven tests added incrementally smaller amounts to the percentage detected).

In the second approach, calculations of the percentage failures detected were separated for baskets (tests 1,2,5, and 6) and paddles (tests 3,4,7, and 8). No single test for either apparatus was able to detect more than 64% of the apparatus-specific failures. However, two combinations of two basket tests had equally high effectiveness by each detecting 82% of the total number of basket failures. For paddles, the best combination of two paddle tests detected 81% of the total number of paddle failures . The two best combinations of two basket tests (tests 1 & 2 or 2 & 6) were combined with the best combination of two paddle tests (tests 3 & 8), and the overall percentages of failures detected were determined. Either combination of four tests detected 86% of the total number of bath failures. Recall that the best overall combination of four tests detected 89% of the failures.

Further increases in percent failures detected for apparatus-specific testing could be realized by combining three basket and three paddle tests

6

Table 3. Percentages of Failures Detected for Combinations of Two Basket Suitability Tests as Compared to Using All Four Basket Tests

Calibrating Condition		1989 Study			1991 Study			1994 Study			Overall		
		No.	Total	%	No.	Total	%	No.	Total	%	No.	Total	%
5	6	2	3	67	4	9	44	5	10	50	11	22	50
1	5	2	3	67	5	9	56	4	10	40	11	22	50
1	6	3	3	100	3	9	33	7	10	70	13	22	59
2	5	1	3	33	7	9	78	7	10	70	15	22	68
1	2	2	3	67	7	9	78	9	10	90	18	22	82
2	6	2	3	67	7	9	78	9	10	90	18	22	82

Table 4. Percentages of Failures Detected for Combinations of Two Paddle Suitability Tests as Compared to Using All Four Paddle Tests

Calibrating Condition		1989 Study			1991 Study			1994 Study			Overall		
		No.	Total	%	No.	Total	%	No.	Total	%	No.	Total	%
3	4	4	6	67	5	8	63	4	12	33	13	26	50
4	8	4	6	67	4	8	50	6	12	50	14	26	54
4	7	5	6	83	5	8	63	5	12	42	15	26	58
7	8	5	6	83	5	8	63	9	12	75	19	26	73
3	7	5	6	83	7	8	88	8	12	67	20	26	77
3	8	5	6	83	7	8	88	9	12	75	21	26	81

7

Table 5. Test or Test Combinations that Detected the Most Failures as Compared to Using All Eight Suitability Tests

Test / Test Combination	Percentage of Failures Detected
2	38 %
2, 3	65 %
2, 3, 8	78 %
2, 3, 7, 8	89 %

(with 95% and 100% of apparatus-specific failures detected, respectively). However, full suitability testing would then involve six tests (three basket plus three paddle tests) or alternatively the four tests identified in Table 5. Neither solution was acceptable, the former had too high a cost for the information gained and the latter overly complicated the testing requirements (i.e., using three different test combinations for basket, paddle, or basket and paddle suitability testing). Thus, we propose using four suitability tests to calibrate a bath for basket and paddle use, and the appropriate two of the four tests for basket or paddle calibration alone.

Recall that there were two equivalent basket combinations involving two tests (1,2 and 2,6; see Table 3). Since the best two test paddle combination (Table 4) includes one salicylic acid and one prednisone tablet, the combination of test 2 and 6 for baskets, which also includes both salicylic acid and prednisone was chosen. Therefore, based on results from the last three collaborative studies, the following dissolution suitability tests were proposed:

To qualify a bath for use with paddles and baskets:

Test Number	Description
2	Prednisone, basket, 100 rpm
6	Salicylic Acid, basket, 100 rpm
3	Prednisone, paddle, 50 rpm
8	Salicylic Acid, paddle, 100 rpm

To qualify a bath for use with baskets only:

Test Number	Description
2	Prednisone, basket, 100 rpm
6	Salicylic Acid, basket, 100 rpm

To qualify a bath for use with paddles only:

Test Number	Description
3	Prednisone, paddle, 50 rpm
8	Salicylic Acid, paddle, 100 rpm

To summarize, the benefits of reducing the number of required suitability tests include reducing the resource required to calibrate a dissolution bath and reducing the consumption of calibrator tablets by half. Extending the life of this final batch (lot L) of prednisone will provide the USP and PhRMA Dissolution Committee more time to identify and implement alternate suitability testing requirements. In addition, the proposed tests detect over 80% of failures, whether calibrating a bath for use with baskets or paddles or both, thus assuring that sufficient control of both apparatuses is maintained.

1996 Collaborative Study Results

The 1996 collaborative dissolution study was completed after the reported analysis to identify a reduced testing scheme. However, once the data from the '96 collaborative study was available, similar analyses to those described above were used to re-evaluate the reduced tests. Overall, in 1996 there were more labs with failing test results, and the failures were more evenly distributed across all test conditions. In other words, there was less distinction between discriminating and non-discriminating suitability tests in the 1996 data. The overall effect of this observation was a reduction in the percent failures detected for all reduced testing schemes. For example, the best combination of two paddles tests (tests 3 and 8) detected only 63% of paddle failures as compared to 83%, 88%, and 75% for the 1989, 1991, and 1994 studies, respectively. Whereas, two basket test combinations (tests 2,6 and 1,2) detected the same percent failures across every previous study, this year the combination of tests 1 and 2 detected 67% of the basket failures, and tests 2 and 6 detected 53% of the basket failures. Combining all data from 1989 through 1996, the combination of tests 1 and 2 detected 76% of the basket failures and tests 2 and 6 detected 70% of the basket failures.

A Recommendation for Reduced Testing...cont.

Despite the distinction between the two basket combinations for the 1996 study results, no changes to the proposed basket combination (tests 2 and 6) are recommended. As described above, the proposed basket test combination of test 2 and 6 is desirable since it uses both salicylic acid and prednisone calibrator tablets.

Looking at the combination of tests to qualify a bath for use with paddles and baskets, the percent failures detected for the previously identified four tests (tests 2,6,3, and 8) decreased to 79% with the 1996 data as compared to 86% for the 1989-1994 data. Note that the best combination of four tests only detected 81% of the failures when the 1996 data was included.

In summary, the results from the 1996 collaborative study are not as strong as the previous years' results in support of the proposed reduced testing scheme. However, given the historical knowledge of the calibrator tablets and dissolution conditions, the proposed reduced testing scheme remains unchanged.

Please direct correspondence to:
Onyenoha Chukwumerije, Ph.D., Chairman
Hoffman-LaRoche, Inc.
340 Kingsland Street
Nutley, NJ 07110-1199
Tel. (201) 235-6904, fax (201) 235-6913
email: onyenoha.chukwumerije@roche.com

References

(1) Lindauer, RL. USP Dissolution Calibration Tablets - Is it Time to Reduce the Amount of Testing Required to Calibrate Apparatus 1 and 2? Pharm Forum 1996; 21(5): 1397-1402.

(2) Qureshi, SA, McGilveray, IJ. A Critical Assessment of the USP Dissolution Apparatus Suitability Test Criteria. Drug Dev and Ind Pharm 1995; 21(8): 905-924.

(3) Skoug, JW. Calibration of Dissolution Rate Apparatuses: A User's Perspective. Dissolution Technologies 1994; 1(2): 3-5.

(4) Workshop Report - AAPS/USP Workshop on Dissolution Calibration and Testing. Pharm Res 1996; 13(1): 6-9.

(5) Cox, DC, Furman, WB, Thornton, LK, Moore, TW, Jefferson, EH. Systematic Error Associated with Apparatus 2 of the USP Dissolution Test III: Limitations of Calibrators and the USP Suitability Test. J Pharm Sciences 1983; 72(8): 910-913.

(6) Moore, TW, Shangraw, RF, Habib, Y. Dissolution Calibrator Tablets: A Recommendation for New Calibrator Tablets to Replace Both Currrent USP Calibrator Tablets. Pharm Forum 1996; 22(3): 2423-2428.

(7) Hamilton, JF, Moore, TW, Kerner, CM. Reproducibility of Dissolution Test Results. Pharm Forum 1995; 21(5): 1383-1386.

(8) Gray, VA, Hubert BB, Krasowski, JA. Calibration of Dissolution Apparatuses 1 and 2 - What To Do When Your Equipment Fails. Pharm Forum 1994; 20(6): 8571-8573.

(9) The USP Dissolution Calibrator Collaborative Study - An Overview of the 1994 Process. Report of the Pharmaceutical Research and Manufacturers of America. Pharm Forum 1994; 20(5):8177-8201.

(10) The 1989 and 1991 test results were provided by David Whiteman (Hoechst Marion Roussel and member of the PhRMA Dissolution Committee). A published source is not available.

(11) Compilation of data into tables of percent failures detected for each test or test combination is automatically completed using the same statistical package used to analyze the collaborative data. Thus, the analysis is based on the statistically-determined acceptable ranges.