

FIGURE 4

OC curves for proposed test compared to OC curve for harmonized pharma copeial test, mean at 96% LC.

test, while the chance to comply with the proposed test is virtually zero.

Although the proposed test is suitable for protection against grossly deviating results from a normal distribution, it is recognized that there could exist hypothesized distributions for which better alternatives exist. Distributions of this nature may indicate within-batch process control issues. Accommodation of all these possible

scenarios is beyond the scope of this discussion, but this could seed further discussion in this area.

CONCLUSION

A test for content uniformity of dosage units that will address the concerns associated with testing large samples using PAT applications has been developed and characterized. The

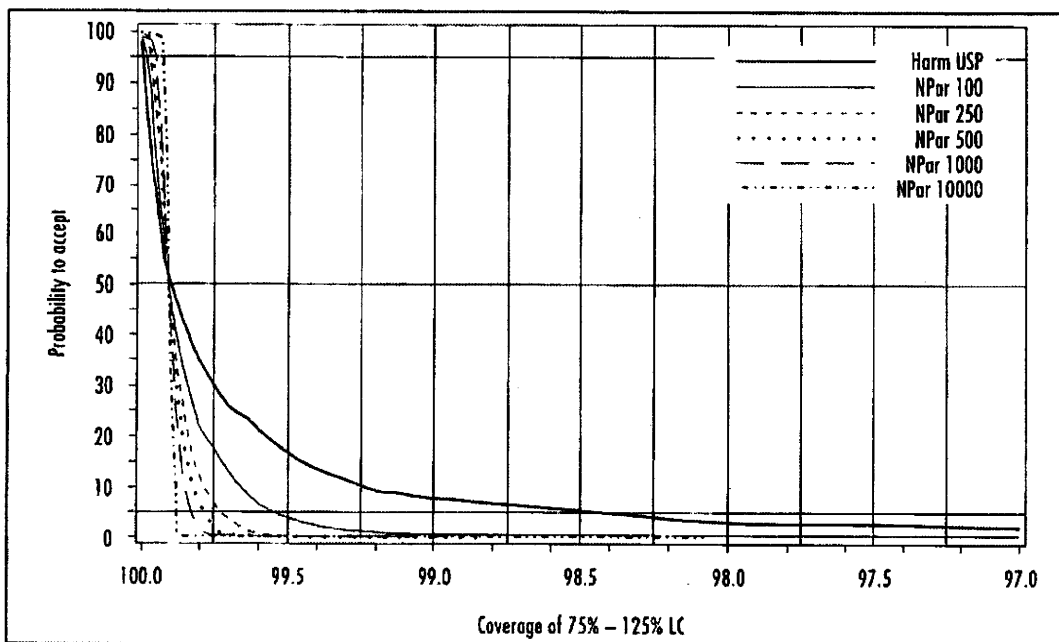


FIGURE 5

Acceptance probability versus coverage of 75% - 125% LC for proposed (different sample sizes) and harmonized pharmacopeial tests (mean at 100% LC).

test is a simple one-tier counting test that provides the required flexibility in selecting sample size. It has been shown that the proposed test

- ensures that the coverage of 85%–115% LC of complying batches (with 95% confidence) is at least 91%,
- guarantees that the acceptance probability of a batch with 95% coverage of 85%–115% LC is at most 50%, and
- provides appropriate control of batches containing grossly deviating results.

In all these aspects the proposed test is an improvement to the harmonized pharmacopeial test for uniformity of dosage units when sample sizes are large. The UDU test proposed in this article can be employed for releasing batches when large sample sizes are collected; when a traditional sample size (10–30 units) is investigated, the harmonized test is more suitable. The proposed test provides one approach to setting batch acceptance criteria for large sample sizes; other tests (including tolerance interval-based tests) may be developed as appropriate.

It is noted that similar issues can be expected for other pharmacopeial tests as the range of PAT applications rapidly grows. These issues can be addressed in a similar fashion after suitable modifications are made, depending on the nature of the issue at hand.

Acknowledgment—We wish to thank Gerry Migliaccio from Pfizer and PhRMA Quality Technical Group for valuable comments on the article.

The Pharmaceutical Research and Manufacturers Association (PhRMA) represents the leading research-based pharmaceutical and biotechnology companies in the United States. PhRMA expert teams and committees are composed of individual specialists from various PhRMA member companies who work as a group to pool their knowledge and expertise and propose concepts and ideas that may result in advances useful for the industry and the public sector.

REFERENCE

- 1 <905>United States Pharmacopeia Convention. Uniformity of dosage units. In: *United States Pharmacopeia*. Rockville, MD: United States Pharmacopeia Convention; 2005:2505–2510.

A PROPOSED CONTENT UNIFORMITY TEST FOR LARGE SAMPLE SIZES

Jim Bergum
Associate Director, Nonclinical Biostatistics
Corresponding Author
Bristol-Myers Squibb Company
1 Squibb Drive
New Brunswick, NJ 08903
james.bergum@bms.com

Kim Erland Vukovinsky,
Director, Nonclinical Statistics
Pfizer
MS 8200-3150;
Eastern Point Rd.
Groton, CT 06340
kim.e.vukovinsky@pfizer.com

A PROPOSED CONTENT UNIFORMITY TEST FOR LARGE SAMPLE SIZES

INTRODUCTION

Applications of Process Analytical Technology (PAT) for in-process and/or end-product release testing are attracting wide interest from both regulators and the pharmaceutical industry. Non-destructive measurement techniques, such as NIR, together with statistical modeling of the obtained spectrum facilitate fast and precise measures supporting the vision of improved manufacturing process understanding. Such techniques can generate a significant amount of data in real time, and thereby open the possibility to improved process control and capability. In such scenarios, the added value of traditional lot release testing based on a random sample from the batch is questioned.

One of the potential applications of PAT is real-time evaluation of tablet content uniformity. A related issue is then the choice of acceptance criteria in light of the increased sample size since the ICH harmonized uniformity of dosage units (UDU) test [1] is based on either 10 or 30 dosage units. To address this issue, the Pharmaceutical Research and Manufacturing (PhRMA) Chemistry and Manufacturing Controls (CMC) Statistics Expert Team (SET) published an alternative to the ICH test [2] in 2006 to facilitate discussion on this topic. Through interactions with European and American regulators during the second half of 2009, questions around the PhRMA approach were raised; this feedback highlighted the need to modify the test to achieve quality equivalent to or better than the ICH UDU test across the entire test range.

This paper provides a modified version of the test provided in the PhRMA CMC SET paper. In both the original and modified tests, the proportion of dosage units within 85-115% of label claim (the "coverage of 85-115% LC") is proposed as the measure of the uniformity of the batch. The acceptance criteria for these tests are based on counting the number of dosage units in the sample outside 85-115% LC and rejecting the batch if that count is too high. The acceptance criteria for these tests are meant to support effective regulatory application of PAT to processing. The tests are nonparametric, simple to implement, use and regulate, and are applicable to large sample sizes. The modified version of the PhRMA CMC SET alternate test provides the same or better assurance as the harmonized UDU test with respect to the batch's uniformity.

PhRMA CMC SET Alternative Uniformity of Dosage Units Test

The EP, JP, USP ICH harmonized pharmacopeial specification (e.g., USP <905> and Ph. Eur. 2.9.40) for the content uniformity of dosage units (UDU) is based on either 10 or 30 dosage units. The PhRMA CMC SET proposed an alternative to the ICH test to address the issue of developing acceptance criteria for increased sample size due to PAT. The tablet sampling procedure (e.g. one tablet every three minutes) is established prior to batch manufacture. The resulting sample size is determined by the sampling procedure, and is larger than 30 dosage units, more likely 100 to 500 units. The PhRMA CMC SET test, called the “Large-N” test in this paper, is a one-tiered counting test for uniformity of dosage units. The nonparametric test is based on counting the number of dosage units outside the 85%-115% of label claim and rejects the batch if that count is outside the set limit.

The number of units (Y) outside the target interval (85-115% LC) is binomially distributed with parameters n (sample size) and p (true proportion outside). When $p = 0.048$, the quality level (QL) calculated corresponds to 95.2% coverage associated with a 50% probability of complying with the ICH-UDU test. The test translates to finding the largest integer t, called c, such that: $c = \max\{t ; \text{Prob}(Y \leq t \parallel p = 0.048) \leq 0.5\}$. Table 1 provides c for various sample sizes. For example, if 500 units are tested and up to (and including) 23 results are outside 85-115% LC, the batch passes the test.

The test is designed to be more conservative than the ICH UDU test for all levels of batch quality where the probability of accepting a batch is less than 0.5. The acceptance limit, c, depends on the sample size, n.

Figure 1 and 2 show OC curves for the Large-N test for sample sizes of 100 and 500 respectively, compared to the OC curve for the ICH-UDU test when the batch mean is 100% LC. Here, the probability to accept the test is plotted against batch standard deviation with curves varying by batch mean. If the mean is off target, the standard deviation must decrease to maintain constant coverage; this results in the shift to the left in the curves. Also, the test is symmetric; a mean of 98% produces the same OC curve as a mean of 102%. Note that for the sample size of 500, the probability of accepting the lot under the large n test is slightly higher than passing the ICH UDU test for small standard deviations; this higher probability is expected since increasing the sample size results in a more discriminating test.

Modified Large-N Test

This paper proposes two modifications to the Large-N paper: (1) reduce the Quality Level (QL) from 0.048 to 0.030 and (2) use the 0.030 QL for all sample sizes from N=100 to N=500. These two modifications result in a more conservative test than the large-N test, since the number of tablets allowed outside of 85%-115% label claim is reduced. The value of 0.030 for the QL was chosen because of its performance against the ICH UDU Operating Characteristic (OC) curve (discussed below) and that the former ICH content uniformity test allowed 1 tablet ($1/30 = 0.033$) outside of the 85%-115% range. Using the same QL for all sample sizes simplifies the calculation for the acceptable number of tablets.

The Modified Large-N is as follows:

- 1) Define the acceptance limit (**c**) by calculating 3.0% of **N** and round it down to the nearest integer. For example, for **N** = 250, 3.0% of the 250 tablets is 7.5, which will be rounded down to **c** =7.
- 2) The batch complies if the number of tablets outside the range of 85.0%-115.0% of label claim (LC) is no more than **c**.

For sample sizes from 100 to 500 tablets, the proposed alternative test is similar to or more conservative than the harmonized compendia test for UDU. A comparison of the acceptance values using Large-N vs. Modified Large-N is provided in Table 2.

Figures 3 - 4 show the OC curves for batch means of 96% and 100% label claim and samples sizes of 100 and 500 for the ICH UDU test as well as the Large-N and Modified Large-N tests. Sample sizes of 100 and 500 were chosen as these cover the current typical range of sample sizes.

The OC curves provide the probability of passing each test at various values of the batch standard deviation (SD). For example, if 100 tablets are tested from a batch with a batch mean of 96% label claim (Figure 3), then for an SD of about 6.4% label claim, the probability of passing either the ICH UDU or Large-N test is about 54% while the probability of passing the Modified Large-N test is about 30%. For the same set of curves at a standard deviation of about 4.0% label claim, the probability of passing either Large-N test or the ICH UDU test is over 99.8%. Also, as can be seen in these figures, the test is sensitive to how far the batch mean is from target (100%). Coverage, or the % outside 85 – 115%, is a function of the average and the standard deviation of the batch. As the batch mean moves away from target, the standard deviation required to achieve the same probability of passing the test is reduced. Therefore, if a batch is

produced off-target the standard deviation needs to decrease to attain constant coverage. The figures also show that the curves become steeper as the sample size increases; this is expected since the test becomes more discriminating as the sample size increases. For sample sizes from 100 to 500, these curves show that the Modified Large-N test is similar to or more stringent than the ICH UDU test.

DISCUSSION ON RESULTS OUTSIDE 75% – 125%

The acceptance criterion for the Large-N and Modified Large-N tests do not include a requirement for zero tolerance of tablets outside 75%-125% LC; which is a requirement of the 30 sample ICH-UDU test. The following provides justification against a zero tolerance rule: there is inherent control of the number of tablets outside 75 – 125% LC through counting the number of tablets outside 85 – 115% and adding a zero tolerance rule would act as a disincentive to collecting the larger sample size which could result in a deterrent to process understanding.

The modified Large-N test inherently controls the number of tablets outside 75 – 125%, through controlling the number outside 85 – 115%. To evaluate the probability of tablets falling outside 75%-125% LC, calculations were performed assuming that the content uniformity results follow a normal distribution. Figure 5 shows the control of tablets outside 75 – 125% through implementation of the Modified Large-N test. Note that for a sample size 500, the probability to accept the batch is approaching 0 when 0.1% of the tablets in the batch are outside 75 – 125% for the modified large-N test whereas the probability of passing the ICH UDU test is greater than 50%.

Calculations were performed for normal distributions that allow a 1%, 2%, or 3% chance of an individual tablet to fall outside 85-115% LC. Table 3 shows the sample size required to detect, with 95% probability that at least one content uniformity result would fall outside 75-125% LC assuming a normal distribution. For a batch mean of 98% LC with a standard deviation of 5.5%, 1% of the results would fall outside 85-115% LC. For this normal distribution, the percent of tablets falling outside 75-125% LC is 0.001473%. A sample size of 203,427 is required to assure with 95% confidence that at least one tablet would be found in the sample. If 10 tablets were sampled per batch, as is the case of the ICH UDU test, then with 95% assurance, at least 1 tablet outside of 75-125% LC would be detected every 20,343 batches. Increasing the sample size from 10 to 500 tablets tested per batch results in at least 1 tablet detected outside 75-125% LC every 407 batches. This is simply the result of the laws of probability; increasing the sample size increases the chance of obtaining a value outside of 75-125% and has no

bearing on the true quality of the batch. The batch quality has not changed, but the batch rejection rate would. Adding a requirement of zero tolerance on tablets outside 75 – 125% would be a disincentive to increasing the sample size and hence a barrier to increased process understanding. Figure 6 shows a plot of the effect of the sample size vs. batch standard deviation.

A requirement that no tablet can fall outside the 75%-125% LC increases the probability of a batch not meeting this limit as the sample size increases. The modified Large-N test only depends on the proportion of tablets outside 85-115% LC, and does not require that the distribution is normal. Table 4 shows the sample size required to detect, with 95% confidence, at least one content uniformity result falling outside 75-125% LC for any distribution.

CONCLUSIONS

The preceding proposes a modified large-N counting test which controls the percent outside 85 – 115% LC to no more than 3%. As the sample size increases (e.g. from 100 to 500) the OC curve becomes steeper, the tests discriminating ability increases, and the level of quality assurance is tightened.

The modified test maintains the beneficial properties of the original test. This test has the advantage of being mathematically simple and simple to implement, requiring only a look-up table or a simple mathematical calculation. There is a-priori flexibility in selecting the sample size. Similar to the original large-N test, this test is non-parametric with the benefit that the test behaves well when the underlying distribution is not normal and testing for normality is not required.

As this proposed test inherently controls the number of tablets outside 75-125% and zero tolerance on this population of tablets could create a disincentive to process understanding, additional test requirements are not recommended.

ACKNOWLEDGEMENT:

Thank you to Fasheng Li, Pfizer for creating the plots used in the paper. Thank you to Tom Garcia, Pfizer and Ambarish Singh, Bristol-Myers Squibb for reviewing the paper.

REFERENCE:

1. USP-NF <905> Uniformity of Dosage Units. In: United States Pharmacopeia. Rockville, MD: United States Pharmacopeia Convention. 2005: 2505–2510.

-
2. PhRMA CMC Statistics Expert Team (Sandell, D.; Vukovinsky, K; Diener, M.; Hofer, J.; Pazdan, J.; Timmermans, J. Development of a Content Uniformity Test Suitable for Large Sample Sizes. *Drug Information Journal* **2006**, 40, 337-344).

Figure 1: Large N (Various %LC) OC Curves for n=100 with ICH UDU (100% LC)

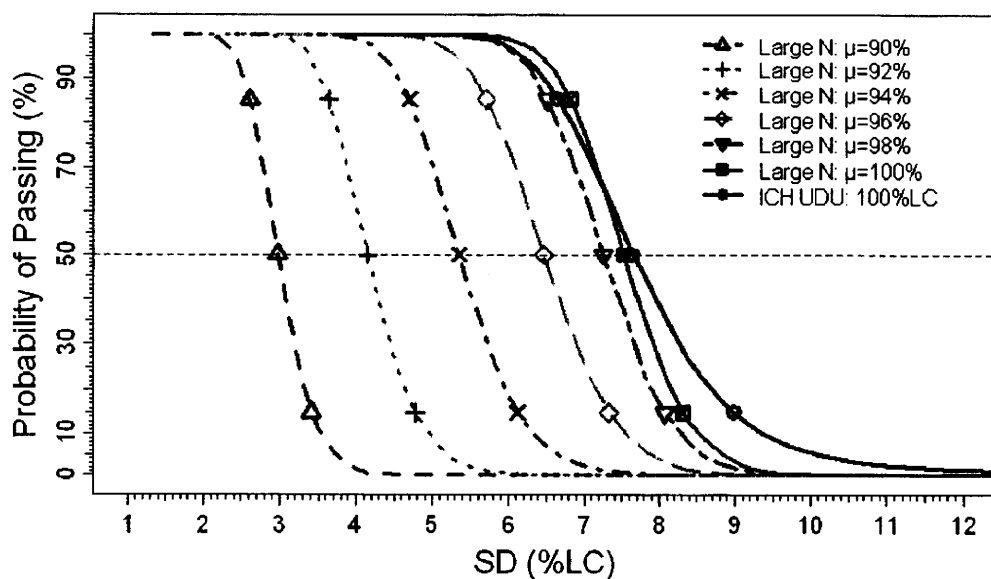


Figure 2: Large N (Various %LC) OC Curves for n=500 with ICH UDU (100% LC)

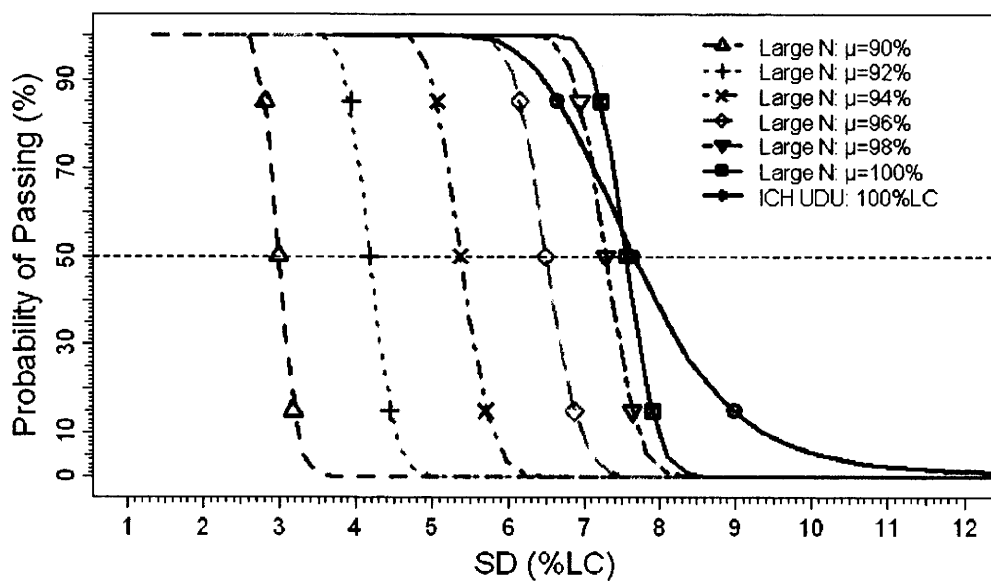


Figure 3: Large N and Modified Large N OC Curves for N=100 with ICH UDU (Batch Mean = 96% and 100% LC)

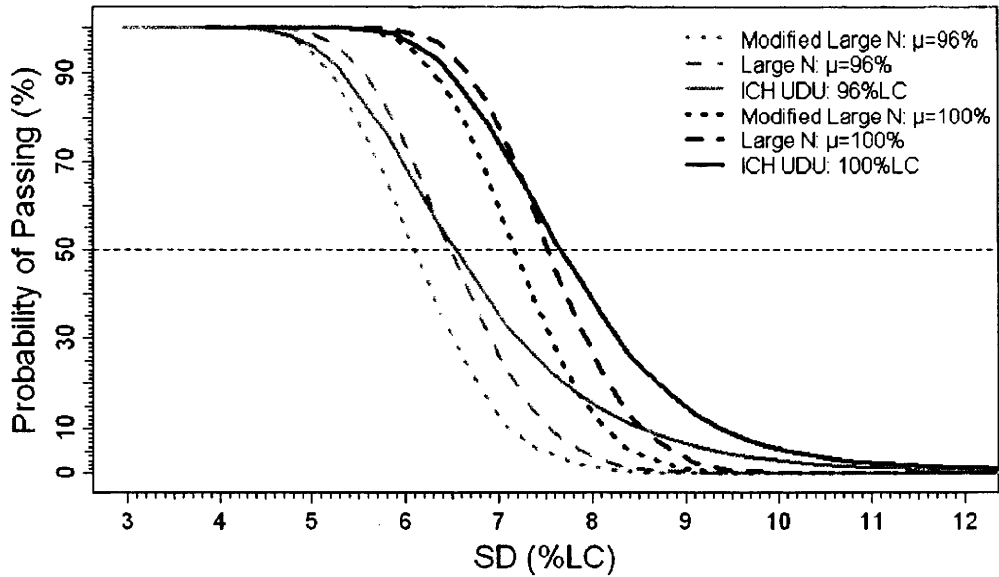


Figure 4: Large N and Modified Large N OC Curves for N=500 with ICH UDU (Batch Mean = 96% and 100% LC)

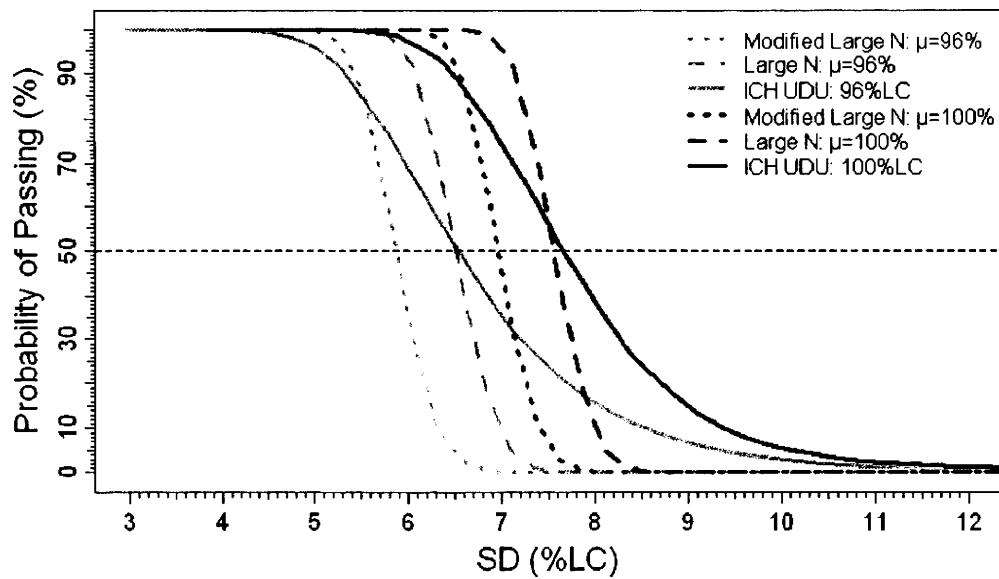


Figure 5: Inherent Control of the Number of Tablets Outside 75 – 125%

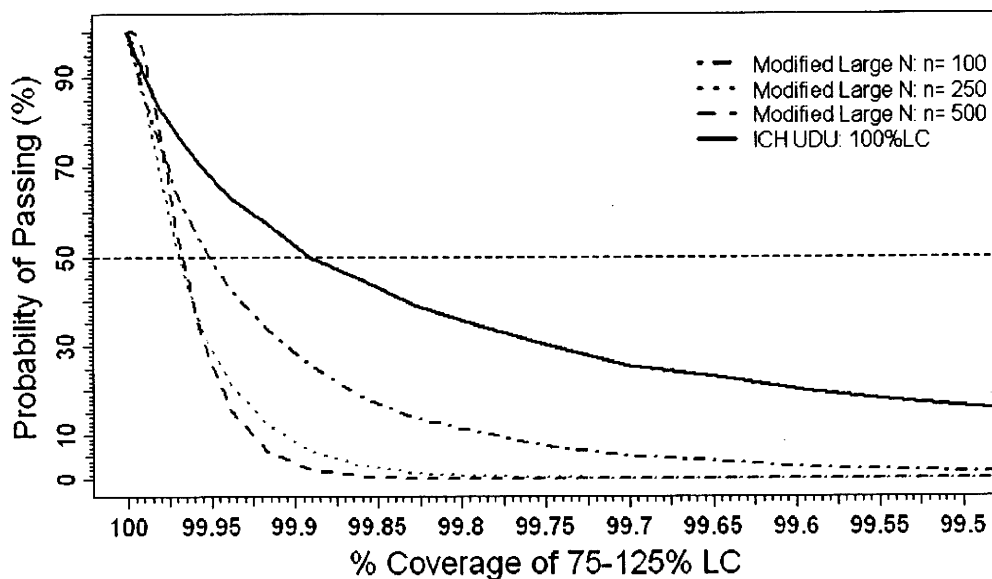


Figure 6: Illustration of Disincentive to Increased Sample Size If Zero Tolerance for Tablets Outside of 75 – 125% (Calculated with Batch Mean 100% LC)

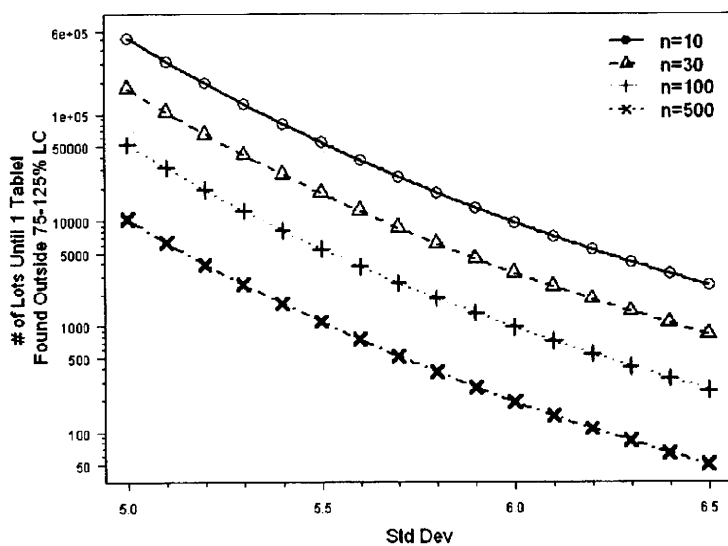


Table 1: Large N Acceptance Limit for Various N

Acceptance Limit, C, for Various Sample Sizes (n)			
n	100	250	500
c	4	11	23

Table 2: Acceptance Values for the Large-N and Modified Large-N Tests

Sample Size	Large-N Acceptance Value	Modified Large-N Acceptance Value
100	4	3
250	11	7
500	23	15

Table 3: Sample Size Required to Detect At Least One Result Outside 75-125% LC With 95% Confidence Assuming Normal Distribution

				Rate: Expected Number of Batches Until One Tablet Outside 75 – 125% For N = 1, 10, 30, 100, 500 sampled per batch				
% Outside 85-115% LC	Batch Mean	Batch Standard Deviation*	% Outside 75-125% LC	1	10	30	100	500
0.5	96	4.27	0.000044	6857735	685774	228592	68578	13716
	98	5.00	0.000215	1394867	139487	46496	13949	2790
	100	5.34	0.000289	1036179	103618	34540	10362	2073
1	96	4.73	0.000443	676302	67631	22544	6764	1353
	98	5.50	0.001473	203427	20343	6781	2035	407
	100	5.82	0.001762	170020	17002	5668	1701	341
2	96	5.35	0.004280	69992	7000	2334	700	140
	98	6.14	0.009620	31140	3114	1039	312	63
	100	6.45	0.010563	28360	2836	946	284	57
3	96	5.82	0.015564	19247	1925	642	193	39
	98	6.62	0.028127	10649	1065	355	107	22
	100	6.91	0.029821	10044	1005	335	101	21

* Batch standard deviation values include all sources of variability (e.g., content uniformity, weight, analytical method, etc.).

Table 4: Sample Size Required to Detect At Least One Result Outside 75-125% LC With 95% Probability For Any Distribution

	Rate: Expected Number of Batches Until One Tablet Outside 75 – 125% For N = 1, 10, 30, 100, 500 sampled per batch					
% Outside 75-125% LC	1	10	30	100	250	500
0.01	29956	2996	999	300	120	60
0.05	5990	599	200	60	24	12
0.1	2994	300	100	30	12	6
0.5	598	60	20	6	3	2
1.0	298	30	10	3	2	1

Enquiry

Demonstration of Uniformity of Dosage Units using Large Sample Sizes

Proposal for a new general chapter in the European Pharmacopoeia

In order to take advantage of increased batch control offered by process analytical technology (PAT) tools, for example large sample sizes used for the control of uniformity of dosage units, a new Ph. Eur. general chapter 2.9.47. Demonstration of uniformity of dosage units using large sample sizes is proposed (see *PA/PH/Exp. PAT/T (11) 1 ANP*, below).

The scientific considerations of the Ph. Eur. PAT Working Party are published in the annex in the form of a scientific article submitted by the rapporteurs for the project.

Please take account of the notice at the start of the section Draft monographs and general texts for comment, on page 297 of this issue of *Pharmeuropa*.

Comments on the proposed general chapter 2.9.47 should be submitted by 30 June 2011.

Reference: PA/PH/EXP. PAT/T (11) 1 ANP

XXXX:20947

2.9.47. DEMONSTRATION OF UNIFORMITY OF DOSAGE UNITS USING LARGE SAMPLE SIZES

Compliance with general chapter 2.9.40. *Uniformity of dosage units* can be demonstrated with the following procedures when the determination is performed on a large sample size ($n > 30$).

Option I (parametric)

SAMPLING PROCEDURE

The consistency of dosage units is evaluated by content uniformity or mass variation as prescribed in Table 2.9.40.-1. Calculate the acceptance value (AV) using the formula and definitions presented in general chapter 2.9.40. Use the acceptability constant (k) defined in Table 2.9.47.-1.

Table 2.9.47.-1. – Acceptability constant applicable to large samples

Sample size (n)	Acceptability constant (k)
≥ 50	1.91
≥ 75	1.87
≥ 100	1.84
≥ 150	1.81
≥ 200	1.79
≥ 300	1.77
≥ 500	1.75
≥ 1000	1.73
≥ 2000	1.72
≥ 5000	1.71
≥ 10 000	1.70

CRITERIA

Apply the following criteria, unless otherwise specified.

The requirements for dosage form uniformity are met if:

- (1) the acceptance value (AV) is less than or equal to $L1$; and
- (2) in the calculation of acceptance value (AV) under content uniformity or under mass variation, the number of individual dosage units ($c2$) with a content less than $(1 - L2 \times 0.01)M$ or more than $(1 + L2 \times 0.01)M$ is within the limit defined in Table 2.9.47.-2.

Table 2.9.47.-2. – Acceptable number of dosage units ($c2$) with a content less than $(1 - L2 \times 0.01)M$ or more than $(1 + L2 \times 0.01)M$

n	< 500	≥ 500	≥ 1000	≥ 2000	≥ 5000	≥ 10 000
$c2$	0	1	2	6	16	34

Option II (non-parametric)

SAMPLING PROCEDURE

The consistency of dosage units is evaluated by content uniformity or mass variation as prescribed in Table 2.9.40.-1. Assay or weigh the units individually. Count the number of units outside the targets $T \pm L1$ ($= c1$) and $T \pm L2$ ($= c2$) respectively, and evaluate if $c1$ and $c2$ are within the limits that are defined in Table 2.9.47.-3.

Unless otherwise specified, $L1$ is 15.0 and $L2$ is 25.0.

Table 2.9.47.-3. – Acceptable number of dosage units outside $T \pm L1$ ($= c1$) and $T \pm L2$ ($= c2$) for a given sample size n .

n	≥ 35	≥ 60	≥ 80	≥ 100	≥ 120	≥ 150	≥ 200
$c1$	1	2	3	4	5	6	8
$c2$	0	0	0	0	0	0	0

n	≥ 300	≥ 500	≥ 1000	≥ 2000	≥ 5000	≥ 10 000
$c1$	13	23	47	95	239	479
$c2$	0	1	2	6	16	34

Annex

Evaluation of Uniformity of Dosage Units using Large Sample Sizes

Ø. Holte¹ and M. Horvat² on behalf of the Ph. Eur. PAT working group

ABSTRACT

Recent developments in analytical technology have made possible the fast determination of unit content in a large number of dosage units from a batch using non-destructive analytical methods. These measurement techniques are often referred to as Process Analytical Technology (PAT). Using such methodology, a better understanding of the manufacturing process and a closer control of the drug product is obtained, as compared with the use of traditional analytical methods. The increased process control that is achieved by PAT is attractive both from the patient's point of view (improved product quality) and from the industry's point of view (increased production efficacy, less batch rejection).

Acceptable batch quality is demonstrated by compliance with the drug product specification. Usually, several of the tests of a specification refer to pharmacopoeial test methodologies and acceptance criteria. One such test is the harmonised Ph. Eur. general chapter 2.9.40. Uniformity of dosage units (UDU). To take full advantage of the increased batch control that is gained by PAT in general and a large sample size in particular, it should be legitimate to demonstrate batch compliance with UDU using other acceptance criteria than those applicable to small samples ($n = 30$). In this paper, a proposal for two new sets of test criteria is presented, applicable to sample sizes larger than 30. The proposal is intended for PAT applications, but is applicable also to traditional analytical methods. The test criteria would be stand-alone alternatives to the current criteria, and it would not be acceptable to re-evaluate a batch against these criteria if it already failed the UDU test for $n = 30$. The sampling plan has to be pre-defined and the sample should be representative of the batch.

1. INTRODUCTION

To ensure the consistency of dosage units, each unit in a batch should have an active substance content within a narrow range around the label claim [1]. Ph. Eur. general chapter 2.9.40. *Uniformity of dosage units* (UDU) is the recommended test to demonstrate this critical property in a batch of drug product. The general chapter was introduced in Supplement 5.2 of the Ph. Eur., and is harmonised with the USP and JP. The earlier Ph. Eur. general chapters 2.9.5. *Uniformity of mass of single-dose preparations* and 2.9.6 *Uniformity of content of single-dose preparations* (content uniformity, CU) are still applicable to existing drug products that were approved before the introduction of UDU. UDU is considered superior to CU in that acceptable and non-acceptable batches, respectively, are more precisely judged using the more recent method. The performance of the old and the new general chapters has been discussed by Limberg and Savsek [2].

(1) Norwegian Medicines Agency

(2) Lek Pharmaceuticals d.d.

It is acknowledged that the evaluation of a small sample will only provide an estimate of the batch quality. There is always a small risk that a highly variable batch would pass the UDU test and be released. Likewise, there is always a risk that a good quality batch would fail the UDU test and be rejected. Increasing the sample size leads to a more precise estimate of the batch variability.

Recently, concern has been raised that the UDU requirements may discourage the use of modern analytical techniques that are fast and non-destructive, often referred to as Process Analytical Technology (PAT) [3-7]. With certain PAT methods, much larger sample sizes are obtained compared to traditional analytical methods. The increased process control that may be achieved when using PAT is attractive both from the patient's point of view (improved product quality) and from the industry's point of view (increased production efficacy, less batch rejection). It is therefore unfortunate that a pharmacopoeial requirement may be regarded as a disincentive to the implementation of such analytical methods.

The main concern with the UDU test when applied to large samples is the requirement that no single result of the test sample is outside $\pm L2$ % of the reference value M ('sample average'; for a precise definition of M , refer to Ph. Eur. general chapter 2.9.40; $L2 = 25.0$, unless otherwise specified). Such an absolute requirement is included in both the CU and the UDU tests. The requirement was established to disclose batches with largely deviating units, even if the sample mean and the overall sample variance are acceptable. This 'safety net' does not assume normal distribution in the sample or in the batch, and it seems reasonable enough not to allow any largely deviating units in a small sample. In our test proposal, a small number of largely deviating units is allowed for large sample sizes.

Another concern with the UDU test is that the improved precision of the batch estimate as a direct result of increasing the sample size beyond $n = 30$ is not taken into account in the evaluation criteria. The UDU test already applies a smaller acceptability constant k to the larger sample size ($k = 2.4$ when $n = 10$; $k = 2.0$ when $n = 30$). Thus, a larger sample standard deviation s is allowed in the calculation of the acceptance value AV when evaluating the larger sample size, provided that sample mean is the same in both samples. In the smaller sample, it is acknowledged that the provided estimate is poor, and a greater security margin is needed to evaluate if the batch is acceptable. In the present proposal for interpretation of the UDU test when applied to large samples, this principle is continued as the sample size increases beyond 30 units.

2. MATERIALS AND METHODS

The test characteristics of the UDU test were analysed in order to establish new test criteria for larger sample sizes. The test characteristics were analysed separately with respect to the calculation of the acceptance value AV and the requirement that no single unit is outside $\pm L2$ % of the reference value M . The new proposals were evaluated by Monte-Carlo simulation of a selection of sample sizes from a selection of batches with relevant properties (normal, bimodal or long-tailed distributions).

2.1. The calculation of the acceptability constant k

Calculation of the acceptability constant k involves the probability that a given proportion of the batch complies with a specified range ($\pm L1$) around the batch average. The statistical parameter is dependent on the sample size, but it is independent of the sample size relative to the batch size, provided that the batch size is still much larger than the sample size. It is however assumed that the sample is representative for the entire batch.

In UDU testing, the required proportion of a batch (coverage, fraction, β) that is within a specified range around the target is determined with a suitable level of probability (confidence level, γ). Assuming that the batch is normally distributed, the prescribed UDU acceptability constants $k = 2.4$ ($n = 10$) and $k = 2.0$ ($n = 30$) imply $\gamma = 84$ % and $\beta = 91$ %. For each sample size, other combinations of γ and β would return k values of 2.4 and 2.0, respectively, but $\gamma = 84$ % and $\beta = 91$ % is the only combination that returns both $k = 2.4$ ($n = 10$) and $k = 2.0$ ($n = 30$). Still assuming that the batch is normally distributed, acceptability constants k can be calculated for various sample sizes n by adopting $\gamma = 0.84$ and $\beta = 0.91$:

$$k(n, \beta, \gamma) = Z_{(1-\beta)/2} \sqrt{\frac{n-1/n}{\chi_{\gamma, n-1}^2}}$$

where $Z_{(1-\beta)/2}$ is the critical value of the normal distribution and $\chi_{\gamma, n-1}^2$ is the critical value of the chi-square distribution with $n-1$ degrees of freedom [8].

When applying $\beta = 0.91$ and $\gamma = 0.84$, k_{∞} converges towards 1.6943 as the sample size n approaches the batch size. In practice, this implies that when the batch mean is on target ($M = \bar{X}$) and $L1 = 15.0$, then the acceptable variability according to UDU, expressed as % RSD (σ), is given by: $AV = L1 = k \cdot s \Rightarrow \sigma = L1/k_{\infty} = 15.0/1.6943 = 8.85$ %. As the sample size decreases, a greater safety margin is required to accept the batch based on the sample variability. Calculated k for a selection of n is presented in Table 1 in section 3.1 below.

2.2. The evaluation of largely deviating units

2.2.1. Acceptance criteria based on the characteristics of the UDU test

For the purpose of detecting the presence of largely deviating units, a non-parametric test is proposed. Following Sandell *et al.* [3], it is possible to calculate the fraction of units outside $\pm L2$ % of the reference value M for a batch that has a 50 % probability to pass the current UDU test. The

probability that all 30 units comply with the requirement is given by $P = (1 - f)^{30}$, where f is the proportion of the batch that is outside $M \pm L2$. Equating P with 0.5 results in $f = 0.02284$. In other words, in a borderline batch where the proportion of units outside $\pm L2$ % of M is 2.284 %, there is a 50 % probability to pass the UDU test with respect to largely deviating units in the sample. In the following, the transition point of a test is defined as the fraction of largely deviating units corresponding to 50 % probability of passing the test. It is assumed that the number of units outside $\pm L2$ % of M in a sample is binomially distributed: $Y \sim \text{Bin}(n, p)$ (sample size n , $p = 0.02284$). Note, however, that no specific assumption is required regarding the distribution of individual dosage units within a batch. Assuming that 2.284 % is the proper acceptability transition point, from the viewpoint of quality, the acceptable number of non-compliant units $c2$ for a given sample size n has been calculated. These results are presented in Table 2 in section 3.2 below. For each sample size n , the cumulative binomial distribution $P(Y \leq c2; n, p)$ defines the largest integer $c2$ that corresponds to a probability to accept of $P \leq 0.5$. An example of the calculation is presented in Figure 1 for sample sizes $n = 250$ and $n = 1000$. From the cumulative distributions it is evident that a 250 or 1000 units sample with 5 and 22 largely deviating units, respectively, would fail the current test criterion for largely deviating units with a probability of slightly less than 50 % ($P \approx 49$), whereas a sample with 6 and 23 largely deviating units, respectively, would fail the test with a probability greater than 50 % (approximately 65 and 57 %, respectively). Consequently, 5 and 22 largely deviating units, respectively, should be acceptable in a 250 or 1000 units sample if the alternative tests should be equally stringent for a borderline batch.

2.2.2. Alternative acceptance criteria with stricter conditions of acceptance

From the clinical efficacy and safety point of view, it may not seem reasonable to allow a batch to contain as much as 2.284 % of units with more than 125 % or less than 75 % of target (when $L2 = 25.0$). Assuming the cut-off is redefined at a point where the current probability to release a batch is $P = 0.75$, then the calculation above yields $f = 9.544 \cdot 10^{-3}$, i.e. 0.954 % of the batch is outside the range. Similarly, a cut-off corresponding to $P = 0.9$ would yield $f = 3.506 \cdot 10^{-3}$, i.e. 0.351 % of the batch is outside the range. The corresponding acceptable number of largely deviating units, $c2$, for a selection of sample size n is presented in Tables 3 and 4 in section 3.2 below.

2.3. Alternative to the calculation of an acceptance value: non-parametric test

Reference is made to Sandell *et al.* [3], where a detailed description of this approach is provided. In a batch that has a 50 % probability to pass the UDU test (the indifference point, transition point), the proportion of units within the acceptable range $\pm L1$ around target (coverage) is 95.2 %. The number of units outside target (Y) in a sample is binomially distributed: $Y \sim \text{Bin}(n, p)$, where n is sample size and $p = (1 - 95.2 \%) = 0.048$. Still assuming the transition point at the coverage of 95.2 %, the acceptable number of non-compliant units $c1$ for a given sample size n has been calculated. For each sample size n , the cumulative binomial distribution $P(Y \leq c1; n, p)$ defines the highest integer $c1$ that corresponds to a probability to accept of $P \leq 0.5$.

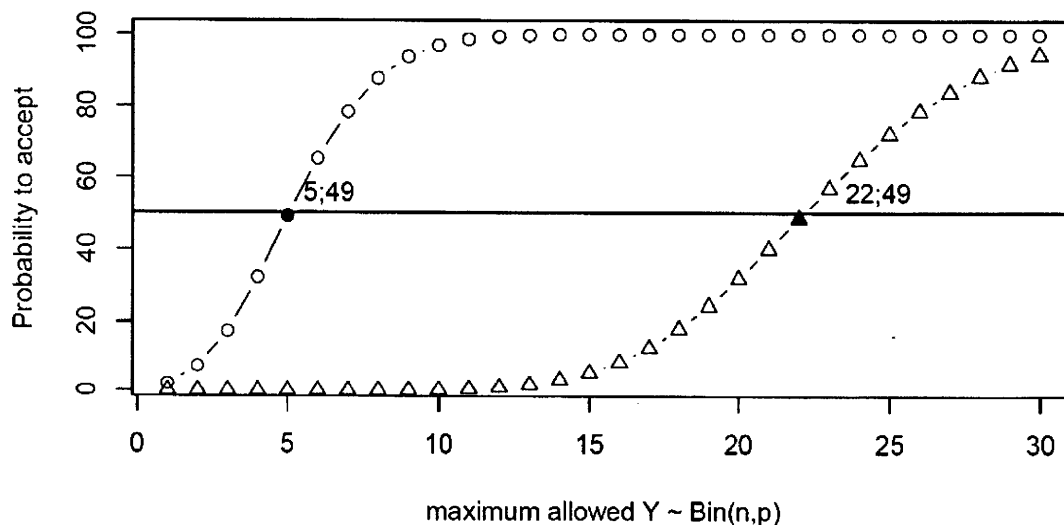


Figure 1 – Example calculation of $c2$ for $n = 250$ (circles) and $n = 1\,000$ (triangles). $p = 0.02284$. The numbers adjacent to the filled points represent the maximum allowed deviating units and the corresponding probability to accept ($\leq 50\%$).

Calculated $c1$ for a selection of n is presented in Table 5 in section 3.3 below.

2.4. Monte-Carlo simulations

The operating characteristic curve (OC curve) for a test demonstrates how the probability to pass the test decreases as the batch variability increases. OC curves for several alternative tests and sample sizes were calculated by a Monte-Carlo simulation. A representative and illustrative selection of OC curves is presented in section 3.4 below to illustrate the performance of the different test methodologies that are discussed in this paper.

In principle, for a specific batch distribution an explicit formula could be provided to calculate this probability. Nevertheless, a direct Monte-Carlo simulation is more robust and versatile in this respect, allowing the straightforward evaluation of any desired batch distribution that is subjected to a certain set of test conditions.

In this procedure, first, a batch distribution is specified. The distribution can be arbitrarily complex, e.g. a mixture of different distributions. Then, a prescribed number of units are randomly generated according to the specified distribution. This randomly generated sample is tested according to the chosen criteria. The cycle of random sample generation and testing is repeated many times (e.g. 10 000 times). Finally, the probability for a batch with such distribution to pass the test is then estimated as the ratio of the total number of cases where the random sample passed the test, to the total number of simulations.

The simulations were performed within the R software environment for statistical computing [9]. Actual source code used will be available upon request.

3. RESULTS AND DISCUSSION

3.1. The calculation of the acceptability constant k

The required proportion of a batch that is within a specified range around the batch average (or target) could

easily generate lengthy discussions among industry and the authorities. What proportion of the batch (coverage, fraction, β) should fall within the acceptable range, and what is a suitable level of probability (confidence level, γ) that the estimate provided by the sample is actually the true batch property? Such discussions are outside the scope of this paper. The authors assume that the test properties (β , γ) that are true for the UDU test should be the same regardless of sample size. Calculated k for a selection of n is presented in Table 1.

Table 1 – Acceptability constant k applicable to large samples

Sample size, n	Acceptability constant, k
≥ 50	1.91
≥ 75	1.87
≥ 100	1.84
≥ 150	1.81
≥ 200	1.79
≥ 300	1.77
≥ 500	1.75
≥ 1000	1.73
≥ 2000	1.72
≥ 5000	1.71
$\geq 10\,000$	1.70

3.2. The evaluation of largely deviating units

The purpose of this requirement is first and foremost to disclose largely deviating units in a batch that is not normally distributed, possibly indicating poor process control. Even in a normally distributed batch, however, there is a small proportion of largely deviating units. When the sample size increases, there is an increasing risk that largely deviating units are included in the sample even for good quality batches. Therefore, the acceptable number of largely deviating units, $c2$, should reflect the sample size.

Table 2 – Acceptable number of dosage units, c_2 , with a content less than $(1 - L_2 \times 0.01)M$ or more than $(1 + L_2 \times 0.01)M$, in samples from a batch that has a 50 % probability to pass the current criteria for $n = 30$ in Ph. Eur. general chapter 2.9.40.

n (exact)	< 75	≥ 75	≥ 117	≥ 161	≥ 292	≥ 467	≥ 993	≥ 2000	≥ 4997	≥ 9992
n (rounded)	< 75	≥ 75	≥ 150	≥ 200	≥ 300	≥ 500	≥ 1000	≥ 2000	≥ 5000	≥ 10 000
c_2	0	1	2	4	6	10	22	45	113	227

Table 3 – c_2 when the transition point is set at where the current probability to release a batch is 75 %.

n (exact)	< 176	≥ 176	≥ 280	≥ 490	≥ 908	≥ 1956	≥ 4995	≥ 9919
n (rounded)	< 200	≥ 200	≥ 300	≥ 500	≥ 1000	≥ 2000	≥ 5000	≥ 10 000
c_2	0	1	2	4	8	18	47	94

Table 4 – c_2 when the transition point is set at where the current probability to release a batch is 90 %.

n (exact)	< 479	≥ 479	≥ 763	≥ 1903	≥ 4754	≥ 9888
n (rounded)	< 500	≥ 500	≥ 1000	≥ 2000	≥ 5000	≥ 10 000
c_2	0	1	2	6	16	34

Calculated c_2 for a selection of n is presented in Table 2. The various n have been rounded in order to establish suitable sets of acceptance criteria ($AV + c_2$, or $c_1 + c_2$ in section 3.3 below). Although the rounding introduces a small error compared to the calculated n , the use of the exact n with three or four significant numbers would leave the impression of a clear statistical basis for the criteria. It is explained in the Introduction that this is not the truth.

Alternative acceptance criteria c_2 for largely deviating units that, within UDU, correspond to borderline batch acceptance probabilities of 75 % and 90 %, respectively, are presented in Tables 3 and 4. It should be recognised that this would represent a tightening of the requirement as compared to UDU. Reference is made to the discussion in section 2.2.2 above.

3.3. Alternative to the calculation of an acceptance value: non-parametric test

In the alternative test procedure, there is an acceptable number of units c_1 for a given sample size n that fall outside the established acceptable range around target. Compared to the original test that was proposed by Sandell *et al.* [3], two changes are however proposed: 1) the test is applicable to all sample sizes $n > 30$, whereas it was initially intended for samples sizes $n \geq 100$ only; and 2) a second test criterion is included to disclose batches with too many largely deviating units. The value of the second test criterion is discussed in section 3.2 above. The proposed test criteria for the largely deviating units are the same as those presented in Table 4

in section 3.2 above. In the test, the assay (or weight) of individual units is determined. The number of units outside target $T \pm L_1 (= c_1)$ and $T \pm L_2 (= c_2)$, respectively, is counted, and the batch is accepted if c_1 and c_2 are within the limits that are defined in Table 5.

3.4. Comparison of the operating characteristics of the current acceptance criteria and the proposed criteria

Typical examples of batch distributions that have been simulated in order to compare the performance of different test criteria are illustrated in Figure 2.

3.4.1. Normally distributed batches

OC curves for normally distributed batches with varying content uniformity, expressed as relative standard deviation, are presented in Figure 3. Each of the batches has been evaluated on increasing sample sizes by three sets of acceptance criteria: in the top panel, the current UDU is compared with the criteria of Tables 1 and 2; in the middle panel, the current UDU is compared with the criteria of Table 1 and the stricter criteria for largely deviating units of Table 4; in the lower panel, the current UDU is compared with the non-parametric test criteria of Table 5. It is evident that the larger the sample size, the more precise is the estimation of the batch variability. This is illustrated by a steeper OC curve. In other words, for a batch with acceptable variability, the probability to pass the test would increase as the sample size increases. Conversely, for a batch with unacceptable variability, the probability to

Table 5 – Acceptable number of dosage units outside $T \pm L_1 (= c_1)$ and $T \pm L_2 (= c_2)$ for a given sample size n .

n (exact)	≥ 35	≥ 56	≥ 77	≥ 97	≥ 118	≥ 139	≥ 181	≥ 285	≥ 493	≥ 993	≥ 1993	≥ 4993	≥ 9993
n (rounded)	≥ 35	≥ 60	≥ 80	≥ 100	≥ 120	≥ 150	≥ 200	≥ 300	≥ 500	≥ 1000	≥ 2000	≥ 5000	≥ 10 000
c_1	1	2	3	4	5	6	8	13	23	47	95	239	479
c_2	0	0	0	0	0	0	0	0	1	2	6	16	34

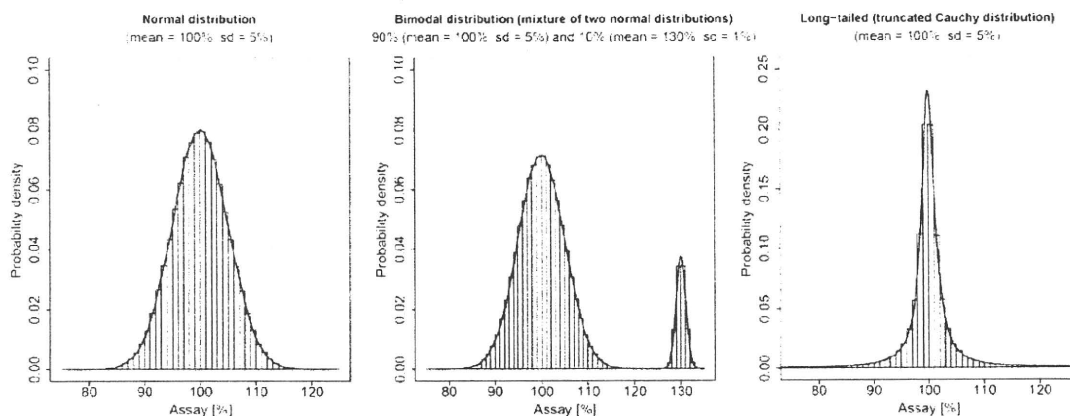


Figure 2 – Typical example of normal distribution (left), bimodal distribution (middle) and long-tailed Cauchy distribution (right).

pass the test would decrease as the sample size increases. It was explained in section 2.1 above, that an on-target batch variability resulting in a RSD of 8.85 % is considered acceptable, according to UDU. Less variability is accepted in batches with off-target mean, but this is not illustrated by the selection of OC curves below. From the middle panel, it is evident that for a normally distributed batch, the *AV* alone is in large responsible for the sample evaluation. However, even in the normally distributed batch, there is a certain proportion of largely deviating units, resulting in a slightly lower probability to pass the test when the more stringent criterion for largely deviating units in Table 4 is applied. From the lower panel, the general picture of the sample size and precision of the estimation of the batch is the same, but the test criteria are somewhat stricter than the acceptance values of Table 1. This could be explained by the different sets of coverage and confidence level that was used in the establishment of the two alternative tests, discussed in sections 2.1 and 2.3 above.

3.4.2. Bimodal batches

OC curves for bimodal batches are presented in Figures 4 and 5. The greater part of the batch is on target (mean = 100 %, RSD = 5 %), but there is a certain proportion of largely deviating units (mean = 130 %, RSD = 1 %). Increasing proportions from 0 to 15 % of largely deviating units are represented in the figures by the abscissa of the OC curve. In Figure 4, each of the batches has been evaluated by the acceptance criteria of the current UDU and the criteria that are described for increasing sample sizes in Tables 1 and 2. The figure demonstrates that the proposed acceptance criteria are much more stringent than the current UDU with respect to disclosing bimodal batches with a significant proportion of largely deviating units. In Figure 5, the increasingly stringent test criteria that are presented in Table 1 and in Tables 2 to 4, and the *AV* criterion alone (Table 1), are compared. It is evident that the bimodal batches are exposed by the alternative criteria for largely deviating units.

3.4.3. Long-tailed distributions

OC curves for batches with a long-tailed Cauchy distribution are presented in Figure 6. Batch variability is represented by increasing standard deviation along the abscissa. Each of the batches has been evaluated by the acceptance criteria that are described in Table 1 and Tables 2 to 4, respectively. The figure demonstrates that a smaller standard deviation is allowed in a batch with a long-tailed distribution compared to a normally distributed batch, due to the greater proportion of largely deviating units in the long-tailed batch (compare to Figure 3 above). Thus, the acceptance value *AV* alone is not suitable to disclose such batches, having an unacceptable proportion of largely deviating units.

4. CONCLUSION

The two alternative UDU tests proposed here should resolve the problems that have been addressed regarding the applicability of the harmonised UDU test when applied to large sample sizes. Moreover, the proposed test criteria are at least equally stringent as the requirements of Ph. Eur. general chapter 2.9.40. In the light of the improved process control that can be achieved when using PAT methodology, the acceptance criteria for largely deviating units presented in Table 4 are proposed, although this would represent a narrowing of the acceptance criteria of the UDU test for sample size $n = 30$. Although the proposed test originally has been motivated by PAT applications, it is applicable also to traditional analytical methods. As Ph. Eur. general chapter 2.9.40 has been harmonised within the Pharmacopoeial Discussion Group (PDG), it is not proposed to change the existing chapter 2.9.40. Instead, it is proposed to publish a separate general chapter in the Ph. Eur. to provide guidance on how to interpret the requirements of Ph. Eur. general chapter 2.9.40 when the sample size is larger than 30 (see draft general chapter 2.9.47. *Demonstration of uniformity of dosage units using large sample sizes*). Depending on the analytical methodology and sample size available, and the purpose of the analysis (e.g. market surveillance on a limited number of samples available or batch release), uniformity of dosage units could be evaluated by a pre-selected set of test criteria: the current criteria for $n = 30$, or any of the proposed criteria for larger sample sizes.