



1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41

A Parametric Tolerance Interval Test for Improved Control of Delivered Dose Uniformity of Orally Inhaled and Nasal Drug Products

Developed and submitted by
The International Pharmaceutical Aerosol Consortium on Regulation and Science (IPAC-RS)
in collaboration with scientists of the Inhalation Technology Focus Group (ITFG) of the
American Association of Pharmaceutical Scientists (AAPS)

In response to the FDA draft Guidances for Industry
*Metered Dose Inhaler (MDI) and Dry Powder Inhaler (DPI) Drug Products
Chemistry, Manufacturing, and Controls Documentation* (Docket No. 98D-0997) and
*Nasal Spray and Inhalation Solution, Suspension, and Spray Drug Products
Chemistry, Manufacturing, and Controls Documentation* (Docket No. 99D-1454)

15 November 2001
FINAL

98D-0997

C 29

41			
42	TABLE OF CONTENTS		Page
43	ABSTRACT		4
44	PROPOSED PARAMETRIC TOLERANCE INTERVAL TEST FOR		
45	DELIVERED DOSE UNIFORMITY TESTING OF OINDP		6
46	PART 1. OVERVIEW OF PROPOSED TEST		8
47	1 Goal.....		8
48	2 Elements of Proposed Test		8
49	3 Proposed Test is a Win-win Solution.....		10
50	PART 2. DETAILS OF PROPOSED TEST		14
51	1 Introduction.....		14
52	2 Desired Properties of DDU Test		15
53	2.1 Quality Definition for Delivered Dose Uniformity.....		15
54	2.2 Consumer and Producer Risks.....		15
55	2.3 Parametric vs Non-Parametric Approaches.....		16
56	2.4 Proposed Limiting Quality.....		17
57	3 Development of the Parametric Tolerance Interval (PTI) Test.....		18
58	3.1 Step 1: Introduction of an Acceptance Value.....		18
59	3.2 Step 2: Extension to Two-tiered Testing.....		21
60	3.3 Step 3: Introduction of Maximum Sample Standard Deviation.....		22
61	3.4 Step 4: Extension to Multiple Test Plans		23
62	3.5 Step 5: Introduction of Requirement on Sample Average.....		27
63	4 Sampling Plans.....		29
64	4.1 Sampling Plan for Single-dose Products		29
65	4.2 Sampling Plan for Multi-dose Products.....		29
66	5 Comparison of PTI Tests with FDA Tests.....		32
67	5.1 Single-dose Products		32
68	5.2 Multi-dose Products.....		33
69	6 Sufficiency of PTI Test Criteria.....		35
70	6.1 The Parametric Criteria of the PTI Test Replaces the Zero-tolerance		
71	Criterion of the FDA DCU Test		35
72	6.2 The Addition of a Zero Tolerance Criterion to the PTI Test would be		
73	Incongruent with the Parametric Approach.....		36
74	6.3 Non-normal Distributions are Treated Conservatively by PTI Tests		37
75	7 Simulated Production Situation		41
76	7.1 Simulated Single-dose Product		41
77	7.2 Simulated Multi-dose Product.....		42

78	CONCLUSIONS.....	45
79	ACKNOWLEDGEMENTS	46
80	APPENDICES.....	47
81	1 FDA Draft Guidance Test	47
82	1.1 Interpretation of the FDA Tests.....	47
83	1.2 Operating Characteristic Curves for the FDA DCU Test.....	48
84	1.3 Operating Characteristic Curves for the Simultaneous Application of the	
85	FDA DCU and TCL Tests	51
86	1.4 Factors Controlling the Outcome of the FDA Tests	53
87	2 Distribution of Data in OINDP Database	56
88	3 Applicability of PTI Test for Non-normally Distributed Data.....	61
89	3.1 Multi-modal Distributions	61
90	3.2 Skewed Distributions	63
91	3.3 Heavy-tailed Distributions	66
92	3.4 Conclusion.....	68
93	4 Simulation and Development of PTI Test Coefficients	69
94	4.1 Algorithm	69
95	4.2 Example (n=10/30)	70
96	4.3 Basis for Simulation	70
97	4.4 SAS® Code	71
98	5 Glossary of Abbreviations, Symbols and Terms	83
99		

100

101

ABSTRACT

102

103

104

105

106

107

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

131

132

133

134

135

136

137

138

139

This paper sets forth an improved delivered dose uniformity (DDU) test for orally inhaled and nasal drug products (OINDP), which is being proposed as a replacement for the uniformity tests (between container and through container life) recommended by the U. S. Food and Drug Administration (FDA) in the following draft Guidances for Industry: (i) *Metered Dose Inhaler (MDI) and Dry Powder Inhaler (DPI) Drug Products* Chemistry, Manufacturing, and Controls Documentation¹; and (ii) *Nasal Spray and Inhalation Solution, Suspension, and Spray Drug Products* Chemistry, Manufacturing, and Controls Documentation².

The test presented here is based on a parametric tolerance interval approach inspired by the work of Dr. Walter Hauck of Thomas Jefferson University, the content uniformity test in the Japanese Pharmacopeia (JP XIII), and by the test in the recently revised Stage 4 draft of Chapter <905> *Uniformity of Dosage Units*³ of the United States Pharmacopeia (USP). A parametric tolerance interval test (PTI test) uses the information obtained from a sample more efficiently than the non-parametric tests recommended in the FDA draft Guidances. This increased efficiency allows the test to provide improved levels of both consumer and producer protection (in the statistical sense) for single-dose products compared to the FDA draft Guidance test. For multi-dose products, the proposed test provides the same high consumer protection as the FDA draft Guidance tests (between container and through container life), while at the same time mitigating the producer risk.

In the proposed PTI test, an 85% coverage of the 75-125% label claim target interval is defined as the default limiting quality standard, below which level there is a low probability of acceptance (<5%). With high confidence, therefore, an accepted batch will have 85% or more of the doses within the specified target interval. These numbers are based on the minimum acceptable quality standards implied by the FDA draft Guidances, the manufacturing capabilities of modern inhalation technology, and the capability of the proposed test.⁴ To ensure the specified batch coverage with at least 95% confidence, the PTI test uses the following criteria on the tested sample:

- An acceptance criterion requiring that the sample standard deviation not exceed a predetermined, sample-size dependent, maximum value.
- An acceptance criterion requiring that an Acceptance Value not exceed a fixed limit. The Acceptance Value is the sum of the absolute deviation of the sample mean from the label claim and the sample standard deviation scaled by a sample-size dependent coefficient.

¹ *Metered Dose Inhaler (MDI) and Dry Powder Inhaler (DPI) Drug Products* Chemistry, Manufacturing, and Controls Documentation, CDER/FDA, October 1998, (Docket No. 98D-0997) <http://www.fda.gov/cder/guidance/2180.pdf>.

² *Nasal Spray and Inhalation Solution, Suspension, and Spray Drug Products* Chemistry, Manufacturing, and Controls Documentation, CDER/FDA, May 1999, (Docket No. 99D-1454) <http://www.fda.gov/cder/guidance/2836.pdf>.

³ *Pharmacopeial Forum* 27(3) p2615.

⁴ For details, see Part 2.

140 An additional acceptance criterion requires the sample mean to be within $100 \pm 15\%$ of label
141 claim (for multi-dose products, this criterion applies to each tested life-stage).

142
143 These criteria serve to control the dose variability and the extent to which the batch and
144 individual doses can deviate from the target. Hence, the distribution of doses for a given batch
145 is well controlled with no need for an absolute limit beyond which no individual sample result
146 is allowed (*i.e.*, no "zero tolerance" limit).

147
148 The proposed test provides several test plans each using a different sample size. An algorithm
149 for calculation of other equally acceptable test plans is provided. All of the test plans ensure
150 the same consumer protection, but have different levels of producer risk. For the producer, this
151 approach provides flexibility in selecting a test plan most appropriate for a particular product,
152 and an incentive to improve product quality (*i.e.*, mean on target and low variability), since
153 with the PTI test, superior product quality is rewarded with the option of reduced testing.

154
155 For multi-dose products, control of through-container-life trends is achieved in the PTI test
156 through a stratified sampling plan, in which one-third of the containers are tested only at the
157 beginning of the container life, one-third only at the middle of the container life and the
158 remaining one-third are tested only at the end of the container life. For products that exhibit no
159 trend or a monotonic trend through container life, testing may be restricted to the beginning and
160 end life stages. The requirements on standard deviation and Acceptance Value are based on the
161 total sample (data from all life stages) whereas the requirement for the mean is imposed for
162 each individual life-stage tested. The proposed stratified sampling plan allows simultaneous
163 control of both between-container and through-container-life uniformity for multi-dose
164 products using a single test.

165
166 IPAC-RS requests that the proposed test replace the tests entitled "Dose Content Uniformity",
167 "Dose Content Uniformity Through Container Life", "Spray Content Uniformity" and "Spray
168 Content Uniformity Through Container Life" in the above-mentioned draft Guidances for
169 Industry. Furthermore, in order to expedite the replacement of the draft Guidance tests, and to
170 facilitate the subsequent public review and comment process, IPAC-RS recommends that the
171 Agency consider issuing a separate draft Guidance for Industry on Delivered Dose Uniformity
172 for Orally Inhaled and Nasal Drug Products. The proposed language for the PTI test for
173 control of DDU in OINDP recommended for inclusion in such a Guidance is contained in the
174 next section of this paper. In Part I of the paper, a general overview of the test is given. Part II
175 provides detailed statistical considerations that support the proposed test. Additional technical
176 details are discussed in the Appendices.

177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220

PROPOSED PARAMETRIC TOLERANCE INTERVAL TEST FOR DELIVERED DOSE UNIFORMITY TESTING OF OINDP

The following test is recommended for adoption by the FDA and industry as a standard procedure for control of delivered dose uniformity of orally inhaled and nasal drug products (OINDP).⁵

Delivered Dose Uniformity

The delivered dose uniformity of orally inhaled and nasal drug products is generally considered acceptable if at least 85% of the doses in a batch fall within $\pm 25\%$ of the delivered dose label claim (LC). In order to claim conformance with this requirement, a confidence level of 95% needs to be demonstrated for the batch. In addition, the sample mean (for each tested life-stage) must be within $\pm 15\%$ of the label claim.

The sponsor should determine a test plan and criteria, consistent with the requirements stated above, that are appropriate for the product in question. Where an auxiliary device is required for the delivery of the preparation from the container, a separate device is typically used for each dose, unless it has been demonstrated that an alternative approach provides equivalent control. The following procedure, using the sample size (n_1 , n_2) and associated acceptability coefficients (k_1 , k_2 , f) of one of the test plans in the *Table of Test Plans* below, ensures, with 95% confidence, at least 85% coverage of the $100\pm 25\%$ LC target interval. These test plans use a two-tiered approach.

For products in single-dose containers (i.e., containers that hold a single individually packaged pre-metered dose unit)

Prepare according to the directions stated in the labeling and measure the amount of drug delivered for n_1 doses. The number of pre-metered units per delivered dose determination should not exceed the number of pre-metered units required for the minimum dose according to the labeling.

For products in multi-dose containers (i.e., containers that hold multiple doses, whether as reservoirs or as ordered assemblies of individually packaged pre-metered dose units)

Prepare according to the directions stated in the labeling and measure the amount of drug delivered for n_1 doses using a separate container for each dose. One-third of the doses are to be sampled from the beginning of container life (first dose after preparation) using $n_1/3$ containers, one-third from the middle of container life (at one-half of the claimed number of deliveries) using another $n_1/3$ containers, and one-third from the end of the claimed number of deliveries using the remaining $n_1/3$ containers. For products that have been demonstrated to exhibit no trend or a monotonic trend through container life, one-half of the doses may be sampled from the beginning of container life using $n_1/2$ containers, and one-half from the end of the claimed

⁵ As explained in detail in Part 2, the limiting quality definition is based on the standards set by the FDA draft Guidances, the manufacturing capabilities of modern inhalation technology, and the capability of the proposed test.

221 number of deliveries using the remaining $n_1/2$ containers. A product is monotonic if the level
222 of the middle life-stage is typically within the range formed by the levels of the beginning and
223 end life stages. The number of deliveries per delivered dose determination should not exceed
224 the number of deliveries in the minimum dose according to the labeling.

225
226 Express the amount of drug delivered for each dose as a percentage of the *delivered dose* label
227 claim. Calculate the Overall Sample Standard Deviation (s), the Overall Sample Mean (m) of
228 the n_1 doses, and the Life Stage Sample Mean (m_{LS}) for each of the life stages tested (note: for
229 preparations in single-dose containers, the Overall Sample Mean and the Life Stage Sample
230 Mean are identical). Accept the batch if:

231 $s \leq 25f/k_1$,
232 $|100-m| + k_1s \leq 25$, and
233 $|100-m_{LS}| \leq 15$ for each life stage tested.

234
235 If not accepted, proceed with the second tier: Observing the directions stated above, measure
236 the amount of drug delivered for n_2-n_1 additional doses to obtain a total sample size of n_2 doses.
237 Express the amount of drug delivered for each dose as a percentage of the delivered dose label
238 claim. Calculate the Overall Sample Standard Deviation (s), the Overall Sample Mean (m) and
239 each of the Life Stage Sample Means (m_{LS}) of the n_2 doses tested. Accept the batch if:

240 $s \leq 25f/k_2$,
241 $|100-m| + k_2s \leq 25$, and
242 $|100-m_{LS}| \leq 15$ for each life stage tested.

243
244 The acceptability coefficients k_1 , k_2 , and f depend on the sample size and several sets of pre-
245 calculated acceptability coefficients are provided in the *Table of Test Plans* below. Other test
246 plans using different sample sizes and/or different number of tiers are acceptable provided that
247 85% coverage of the target interval is ensured with 95% confidence.

248
249 For products where safety and/or efficacy concerns indicate a need for a higher level of
250 uniformity, tighter limits on the coverage and/or target interval may be warranted. If there is
251 adequate clinical evidence to support a lower level of uniformity, less stringent requirements on
252 the coverage and/or target interval may be acceptable.

253

254 **Table of Test Plans.** Details of 6 two-tiered test plans giving 95% confidence of at least 85%
255 coverage of the target interval ($100 \pm 25\%$ LC). See text for application.

Test plan	Sample Size at Tier 1	Total Sample size Tier 1+2	Acceptability Coefficient at Tier 1	Acceptability Coefficient at Tier 2	Factor for Maximum Sample Standard Deviation
	n_1	n_2	k_1	k_2	f
1	10	30	2.09	1.59	0.839
2	12	36	1.95	1.52	0.826
3	14	42	1.85	1.48	0.819
4	15	45	1.81	1.46	0.815
5	18	54	1.72	1.42	0.808
6	24	72	1.59	1.36	0.796

256

256

257

PART 1. OVERVIEW OF PROPOSED TEST

258

259

260

1 Goal

261

262

263

264

265

266

267

268

269

270

271

272

273

274

275

276

277

278

279

280

281

282

283

284

285

After careful review of the FDA draft Guidances for Industry and an assessment of an extensive industry data-base on delivered dose uniformity (DDU),⁶ the Dose Content Uniformity (DCU) Working Group of the IITFG/IPAC-RS Collaboration set a goal of developing a DDU test that would accomplish the following, compared to the FDA draft Guidance tests⁷:

- Improved ability to characterize batch quality;
- Same or improved consumer protection; and
- Improved producer protection.

In general, a DDU test should control the mean delivered dose for the batch as well as the variability of delivered doses in the batch between different containers and, for multi-dose preparations, within containers (including through-container-life trends). The control should be relative to the label claim delivery.

Ideally, the desired DDU test should ensure a consistent minimum quality standard for a wide variety of orally inhaled and nasal drug products. In addition, it would be advantageous for producers and reviewers if the same test and corresponding criteria could apply to a wide variety of testing situations, such as routine release testing, validation, stability and investigational studies.

All of the objectives outlined above are accomplished by the proposed Parametric Tolerance Interval test (PTI test).

286

2 Elements of Proposed Test

287

288

289

290

291

292

293

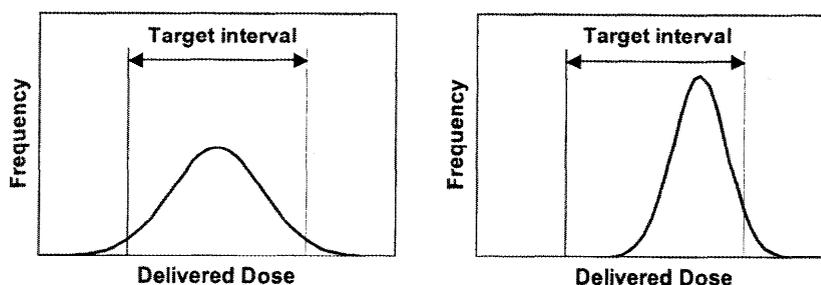
As a measure of batch quality, the PTI test uses *coverage*, or the proportion of doses that fall within a specified target interval. Graphically, coverage represents the area under the distribution curve within a given target interval (see Figure A). For example, the two distributions shown in Figure A have equal coverage of the indicated target interval, and thus they are of equally acceptable quality.

Figure A further illustrates the following features of using coverage as a measure of quality:

⁶ *Initial Assessment of the IITFG/IPAC Dose Content Uniformity Database by the CMC Specifications Technical Team of the IITFG/IPAC Collaboration* (July 2000), available at http://www.fda.gov/ohrms/dockets/ac/00/techrepro/3609_reports.htm.

⁷ *Metered Dose Inhaler (MDI) and Dry Powder Inhaler (DPI) Drug Products* Chemistry, Manufacturing, and Controls Documentation (Docket No. 98D-0997) <http://www.fda.gov/cder/guidance/2180.pdf>, and *Nasal Spray and Inhalation Solution, Suspension, and Spray Drug Products* Chemistry, Manufacturing, and Controls Documentation (Docket No. 99D-1454) <http://www.fda.gov/cder/guidance/2836.pdf>.

- 294 • The concept of coverage allows one to express quality in a standard and consistent manner
 295 while focusing on the target (label claim delivery).
- 296 • The requirement of minimum coverage allows a trade-off between the mean and variance
 297 (*i.e.*, as the mean dose is drifting off target, the standard deviation needs to become tighter
 298 in order to surpass the minimum coverage).
- 299



300 **Figure A. Graphs illustrating coverage, which is the proportion of doses in the batch that are**
 301 **within a specified target interval. The two distributions represented in this figure**
 302 **have equal coverage of the indicated target interval.**
 303
 304
 305

306 In the proposed PTI test, an 85% coverage of the 75-125% label claim (LC) target interval is
 307 defined as the minimum quality standard⁸, below which level there is a low probability of
 308 acceptance (<5%). To ensure the specified coverage of each accepted batch with at least 95%
 309 confidence, the PTI test uses the following parametric criteria for the sample:

- 310 • An acceptance criterion requiring that the sample standard deviation not exceed a
 311 sample-size dependent maximum value. This requirement controls the batch variability
 312 when the mean is close to the label claim.
- 313 • An acceptance criterion requiring that an Acceptance Value not exceed a fixed limit.
 314 The Acceptance Value is the sum of the absolute deviation of the sample mean from the
 315 label claim and the sample standard deviation scaled by a sample-size dependent
 316 coefficient. The Acceptance Value simultaneously controls the batch mean and
 317 standard deviation so that less variability is allowed the more the mean deviates from
 318 the label claim.
- 319

320 Together, these requirements control the coverage and the extent to which individual values
 321 may deviate from the label claim.

322

323 In addition, an acceptance criterion of 85-115% LC is imposed on the sample mean to further
 324 control mean deviation from the label claim when variability is low (for a multi-dose product,
 325 this criterion is applied to each tested life-stage, *e.g.*, beginning, middle and end of container
 326 life).

327

⁸ As explained in detail in Part 2, these numbers are based on the standards set by the FDA draft Guidances, the manufacturing capabilities of modern inhalation technology, and the capability of the proposed PTI test.

328 In practice, a PTI test would involve the following steps (for a single-dose product):

329

330 **MEASURE** a pre-defined number of doses, n_1

331

332 **CALCULATE** (1) sample mean, m , expressed in % LC

333

(2) sample standard deviation, s , expressed in % LC, and

334

(3) Acceptance Value⁹ $AV = \sqrt{100-m} + ks$

335

336 **COMPARE:** (1) $\sqrt{100-m} \leq 15$;

337

(2) $s \leq \text{Maximum Sample Standard Deviation (MSSD)}^{10}$;

338

(3) $AV \leq 25$

339

340 *2nd Tier: If these criteria are not met in the first tier, second tier testing*

341

is performed. In the second tier, the steps are repeated with a larger

342

sample size and an adjusted coefficient k .

343

344 For multi-dose products, control of through-container-life trends is achieved in the PTI test
345 through a stratified sampling plan that captures both inter- and intra-container variability. One-
346 third of the containers are tested in the beginning, one-third in the middle, and one-third in the
347 end of the labeled number of deliveries (if justified for the product, testing the middle life stage
348 may be waived¹¹). Additional control over container through-life trends is achieved by
349 applying the criterion for the mean separately to each of the tested life stages.

350

351 Using this design, the information for multi-dose products sought by the FDA through two
352 uniformity tests in the draft Guidances, is captured in a single test.

353

354

355 **3 Proposed Test is a Win-win Solution**

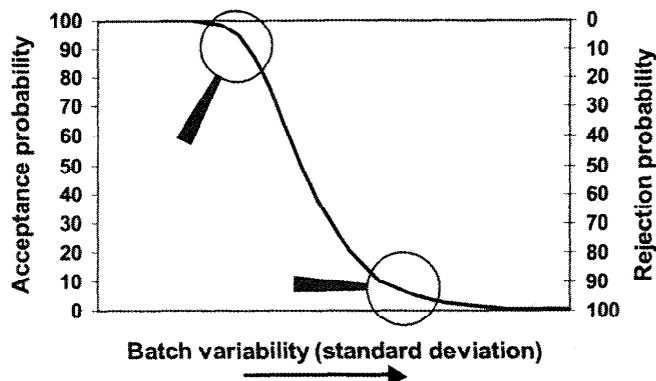
356 In order to compare different tests (*e.g.*, FDA and PTI), one has to analyze the operating
357 characteristics of the tests in question, which is typically accomplished by using computer
358 simulations. A conventional way of describing operating characteristics of a test is to plot the
359 probability of acceptance as a function of a quality parameter, such as the batch standard
360 deviation (see Figure B). The resulting curve is commonly known as an operating
361 characteristic (OC) curve. For a batch with a given mean, the acceptance probability decreases
362 as the batch standard deviation increases.

363

⁹ The sample-size dependent coefficient k is found in the *Table of Test Plans*, page 7.

¹⁰ MSSD is defined as $25f/k$. The sample-size dependent coefficients f and k are found in the *Table of Test Plans*, page 7.

¹¹ See Part 2, Section 4.2.1.



364
365 **Figure B. An operating characteristic (OC) curve.**
366

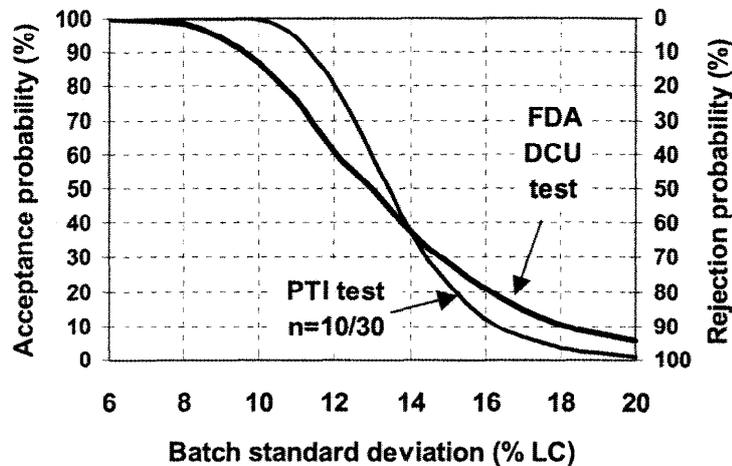
367 The bottom-right part of an OC curve refers to batches of high variability. Note that due to
368 sampling variability, the corresponding probability of acceptance is small but not zero. Thus,
369 this region of the curve represents the consumer risk that a batch of sub-standard quality may
370 be accepted due to pure chance. For example, a 95% probability of rejection means that there
371 is a 5% probability of accepting a batch of high variability (low uniformity).
372

373 The top-left portion of the curve represents batches having a low standard deviation. Due to
374 sampling variability, the probability to accept such batches is less than 100%, and this region of
375 the curve represents the producer risk that a batch of acceptable quality may nevertheless be
376 rejected due to pure chance. For example, a 95% probability of acceptance means that there is
377 a 5% probability of rejecting a batch of low variability (high uniformity). The producer and
378 consumer risks defined in this way are also known as probabilities of Type II and Type I errors,
379 respectively.
380

381 One way to evaluate different tests is to compare the batch quality corresponding to a certain
382 consumer risk. If the quality of batches released with 5 % probability is improved, then such a
383 test has improved consumer protection.
384

385 Figure C shows a comparison of the operating characteristic curves pertaining to single-dose
386 products with batch means on target for the FDA dose content uniformity (between container)
387 test (denoted here as the FDA DCU test¹²) and the PTI test using the same n=10/30 sampling
388 plan (10 observations in 1st tier, total of 30 observations after 2nd tier). As one can see, the OC
389 curve for the PTI test lies *below* the OC curve for the FDA DCU test in the region of high
390 standard deviations. Thus, the proposed test will more likely reject batches of poor uniformity
391 (high variability) compared to the FDA DCU test, *i.e.*, the consumer risk is reduced. Looking
392 at it another way, the quality of batches accepted with 5% probability is improved with the PTI
393 test compared to the FDA test. In other words, consumer protection is improved.
394

¹² See this entry in the Glossary for the exact reference.

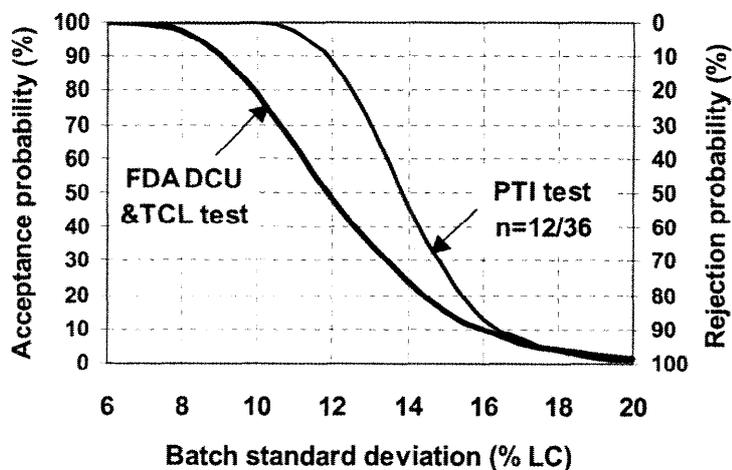


395
396
397
398

Figure C. Operating characteristic curves for the FDA DCU test and the PTI test using the $n=10/30$ sampling plan for single-dose products with mean at target.

399
400
401
402
403
404
405
406
407

Figure D shows a comparison of the operating characteristic curves pertaining to multi-dose products with batch means on target for the combined application of the FDA dose content uniformity and dose content uniformity through-container-life tests (denoted here as the FDA DCU&TCL test¹³), and the PTI test using a $n=12/36$ sampling plan. As can be seen, the OC curves for the FDA DCU&TCL test and the PTI test intersect in the region of high standard deviations. Thus, the proposed PTI test will reject batches of poor uniformity (high variability) with similar confidence as the FDA DCU&TCL test, *i.e.*, the consumer protection is comparable.



408
409
410
411
412

Figure D. Operating characteristic curves for the simultaneous application of the FDA DCU and TCL test, and the PTI test using a $12/36$ sampling plan, for multi-dose products with mean at target.

¹³ See this entry in the Glossary for the exact reference. Also see Appendix 1, Section 1.3.

413 On the other hand, in the region of low standard deviations, the PTI test curve lies *above* that of
414 the FDA test, both for the tests for single-dose products (Figure C) and for multi-dose products
415 (Figure D). This means that the producer will have a higher probability to pass batches of high
416 uniformity (low variability) compared to the FDA draft Guidance tests. In other words, the
417 producer risk is reduced, which creates the win-win solution.

418
419 The reason that the consumer protection is *improved* with the PTI test compared to the test
420 recommended in the FDA draft Guidances for single-dose products, while on the other hand,
421 the tests have *comparable* consumer protection for multi-dose products, is due to the fact that
422 the draft Guidances recommend a more stringent test for multi-dose products than for single-
423 dose products. The proposed PTI test provides equal consumer protection for both types of
424 products, and at the same high level as that implied by the draft Guidances for multi-dose
425 products.

426
427 The proposed test provides several test plans using different sample sizes, starting from
428 $n=10/30$. All of the test plans ensure the same consumer protection, but have different levels of
429 producer risk. This provides an incentive for the industry to improve product quality, since
430 with the PTI test, superior product quality is rewarded with the option of selecting a plan with a
431 reduced sample size. The details of six two-tiered sample plans are listed in the *Table of Test*
432 *Plans* (page 7). If a different plan is desired, Appendix 4 provides the algorithm to calculate
433 acceptability coefficients assuring the same consumer protection for the preferred choice of
434 sample size.

435
436 In summary, the main features of the proposed test are the following:

- 437 • A parametric tolerance interval test is proposed to replace the non-parametric tests in
438 the FDA draft Guidances. The proposed PTI test uses the information obtained from a
439 sample more efficiently.
- 440 • The proposed test explicitly defines batch quality in terms of the minimum proportion
441 of doses within the 75-125% LC target interval (*i.e.*, 85% coverage or more) and
442 requires this to be ensured with high confidence (*i.e.*, 95% probability) for each batch.
- 443 • For single-dose products, the parametric test simultaneously reduces consumer and
444 producer risks.
- 445 • For multi-dose products, the consumer protection is maintained at the high level
446 recommended in the FDA draft Guidances with a simultaneous reduction in the
447 producer risk.
- 448 • The parametric test comprises a number of test plans using different sample sizes, each
449 providing equivalent consumer protection. Improved quality is rewarded with a
450 lowered producer risk or with the option of selecting a reduced sample plan.

451

452

453

PART 2. DETAILS OF PROPOSED TEST

454

455

456

1 Introduction

457

458

459

460

461

462

463

464

465

466

467

468

469

470

471

472

473

474

475

476

477

478

479

480

481

482

483

484

A properly designed test for delivered dose uniformity of orally inhaled and nasal drug products must be applicable to all such products regardless of particular therapeutic indication or delivery device (*i.e.* MDI, DPI, single-dose, multi-dose, sprays, *etc.*).

On the other hand, it is impossible to identify a universal level at which clinical safety or efficacy would be compromised because clinical considerations vary based upon active ingredient and therapeutic indication.

Therefore, in this proposal dose uniformity is treated purely as a quality issue. Thus, an explicit statement of limiting batch quality is the most critical element of a properly designed delivered dose uniformity test. A test may then be designed to ensure with high confidence that a batch at or below the limiting quality is not accepted, thereby protecting the consumer from sub-standard quality batches. There is no intention to imply that this quality standard is generally required for safety or efficacy reasons.

The present proposal sets forth both the test that assesses batch quality with high confidence, and the limiting quality statement that reflects the quality standards required by the FDA and the manufacturing capabilities of modern inhalation technology. Furthermore, this proposal provides multiple test plans using different sample sizes, each satisfying the limiting quality statement, so that the producer may select the sample size most appropriate for the product in question. Since all test plans provide the same limiting quality, consumer protection is not affected.

In the final analysis, the specification for a particular drug product could be affected by clinical evidence as well. This could potentially result in a specification that is either more or less strict than that contained in this proposal.

484

485 2 Desired Properties of DDU Test

486 2.1 Quality Definition for Delivered Dose Uniformity

487 The primary purpose of a DDU test is to control dose uniformity of a batch. It would also be
488 advantageous if the same principles and corresponding criteria could be employed, for
489 example, in product development, stability investigations and process validations. In practice,
490 dose uniformity of a batch must be judged based on the properties of a finite representative
491 sample from the batch.

492

493 The fundamental starting point for being able to develop suitable requirements is to define an
494 end-point (*i.e.*, metric) that describes what is meant by “quality” for DDU. When this has been
495 done, one needs a limit, which, in terms of the selected quality end-point, defines what quality
496 should be considered minimally acceptable (*i.e.* the limiting quality). Once this is
497 accomplished, a statistical determination is possible of how sample information is best utilized
498 to decide on the disposition of a batch.

499

500 As stated above, the first step in defining the limiting quality is to decide on the metric to be
501 used for characterizing uniformity. Following Dr. W. Hauck’s approach, this has been defined
502 in terms of two factors, the target interval and the coverage of the target interval (the proportion
503 of doses in the batch that are within the interval). The coverage of a target interval is an
504 appropriate standard metric of quality, as uniformity around a fixed target (delivered dose label
505 claim) is the ultimate goal. The next step in defining the limiting quality is to decide on
506 quantitative limits for the metrics. The quality criterion in this case is the width of the target
507 interval and the true proportion (or coverage) of dose values that fall within this interval
508 (referring to the batch, not the sample). An example of a limiting quality definition is “not less
509 than 85% of the doses in a batch fall within the interval $100 \pm 25\%$ of the label claim”.

510

511 By an appropriate choice of sample acceptance criteria, tests may be devised which provide a
512 correlation between the actual coverage of the target interval and the probability that a sample
513 from the batch will comply with these criteria. This is a desired characteristic of a test, because
514 it provides a transparent link between batch quality and the probability of complying with the
515 test.

516

517 2.2 Consumer and Producer Risks

518 Once the limiting batch quality has been defined, the hypothesis that the batch fulfills the
519 defined quality criteria can be tested – at the desired level of confidence – by inspection of a
520 sample from the batch. In developing the proposed test, the generally accepted confidence
521 level of 95% has been used.

522

523 By back-calculation it is possible to determine sample criteria that will ensure with 95%
524 confidence that an accepted batch fulfills the limiting quality criteria. That is, a sample
525 conforming to such sample criteria ensures, with a 5% risk of error, that the batch quality is
526 equal or superior to the limiting quality. In other words, the limiting quality is the quality at
527 which an isolated test of a sample from the batch has a low (5%) probability of acceptance.

528 Therefore, the consumer is facing a 5% risk that a batch at the limiting quality will be accepted.
529 This consumer risk of a false acceptance is alternatively known as the risk of a Type I error.

530

531 In addition to protecting the consumer from sub-standard quality batches, the producer should
532 be protected from the risk of rejecting acceptable batches. From the sample criteria, it is
533 possible to calculate the quality of a batch that would be accepted with 95% probability. For
534 batches with this quality, or better, the probability that a sample would *not* pass the criteria is at
535 most 5%. This is called the producer risk of a false rejection, or alternatively, the risk of a
536 Type II error.

537

538 The quality at which the batch has a 95% acceptance probability (5% producer risk) is, of
539 course, better than the quality at which it has a 5% acceptance probability (5% consumer risk).
540 This fact only serves to further protect the consumer, as it is in the producer's best interest to
541 manufacture batches with a quality that ensures at least 95% probability of acceptance.

542

543 The magnitude of the difference between the quality at 5% and 95% acceptance probabilities is
544 determined primarily by the following three factors:

545

- the definition of the limiting batch quality,
- the sample size, and
- the efficiency of the test, *i.e.*, the ability of the test to extract and use the information
548 obtained from a sample to characterize the batch.

549

550 2.3 Parametric vs Non-Parametric Approaches

551 Both non-parametric and parametric approaches can be used to test for coverage of a target
552 interval. The dose uniformity tests in the FDA draft Guidances are non-parametric (as are most
553 commonly used uniformity tests for pharmaceutical products). The proposed test is parametric.

554

555 A non-parametric test does not presuppose any particular distribution. The major attribute the
556 non-parametric FDA uniformity tests use is whether an observation is within or outside a target
557 interval. In this test, only the count of observations falling within a fixed interval is used to
558 estimate the population coverage. In cases where a distributional assumption is reasonable, a
559 non-parametric method is not the most efficient approach.

560

561 Assuming that the data do follow a normal distribution, the sample mean and standard
562 deviation are sufficient statistics to characterize the batch. This means that the information in
563 the sample can be summarized by the sample mean and standard deviation *without* any loss of
564 information. Further, as these statistical parameters can be used to estimate the batch coverage,
565 limits can be found for the mean and standard deviation that assure (at the selected level of
566 confidence) that the batch coverage is not less than the limiting quality. A parametric test
567 based on a sufficient statistic provides a more complete and thorough use of the data, and
568 therefore provides a more precise estimate of quality, compared to a non-parametric test using
569 the same sample size.

570

571 The assumption of normality used for developing the proposed test can be investigated by
572 examination of actual data when a large number of observations are available. The database of
573 OINDP collected by ITFG/IPAC-RS contains 46,816 results for 80 products that demonstrate
574 that the normality assumption is very reasonable (Appendix 2). At this juncture it will be
575 assumed that data for all products are normally distributed. However, one of the desired

576 properties of a parametric delivered dose uniformity test must be that consumer protection is
577 not eroded if the normality assumption is violated. Section 6.3 and Appendix 3 provide
578 evidence that the proposed parametric test is more conservative in its treatment of non-
579 normally distributed data as compared to normally distributed data. Thus, the proposed test
580 does not compromise consumer protection in cases when data are non-normally distributed.
581

582 **2.4 Proposed Limiting Quality**

583 The target interval selected for the proposed test is $100\pm 25\%$ LC. This target interval was
584 chosen because the FDA draft Guidance tests use the $100\pm 25\%$ LC interval for the outer limits
585 criterion, the criterion that predominantly determines the outcome of the FDA uniformity tests
586 (see Appendix 1).
587

588 The coverage of the target interval that defines the limiting batch quality in the proposed test
589 was set to 85%. This figure resulted from the careful consideration of the following factors:

- 590 • When expressed in terms of coverage of the 75-125% LC target interval, the limiting
591 quality implied by the FDA DCU test¹⁴ for single-dose products, as determined at the
592 95% confidence level, is 78%¹⁵, which is a lower coverage than provided by the
593 proposed PTI test;
- 594 • When expressed in terms of coverage of the 75-125% LC target interval, the limiting
595 quality implied by the combined application of the FDA DCU&TCL tests¹⁶ for multi-
596 dose products, as determined at the 95% confidence level, is 85%¹⁷, which is equal to
597 the coverage provided by the proposed PTI test; and
- 598 • 85% coverage does not require an unreasonable amount of testing to achieve 95 %
599 confidence given the current capabilities of inhalation technology and the capabilities of
600 the proposed test. Higher levels of coverage would require significantly more testing to
601 achieve 95% confidence.
602

¹⁴ See this entry in the Glossary for the exact reference. The FDA DCU test is recommended by the draft Guidances for all products.

¹⁵ The draft Guidances do not provide an explicit quality statement regarding uniformity. Therefore, the operating characteristics of the DCU test were determined prior to the development of the present proposal (see Appendix 1, section 1.2). This “reverse engineering” approach was used to determine that the limiting quality that is implied by the FDA DCU test is equivalent to 78% coverage of the $100\pm 25\%$ LC interval.

¹⁶ See this entry in the Glossary for the exact reference. The FDA TCL test is recommended by the draft Guidances in addition to the FDA DCU test for all multi-dose products.

¹⁷ The draft Guidances do not provide an explicit quality statement regarding uniformity. Therefore, the operating characteristics of the simultaneous application of the DCU and TCL tests were determined prior to the development of the present proposal (see Appendix 1, section 1.3). This “reverse engineering” approach was used to determine that the limiting quality that is implied by the FDA DCU&TCL test is equivalent to 85% coverage of the $100\pm 25\%$ LC interval.

602

603 **3 Development of the Parametric Tolerance Interval** 604 **(PTI) Test**

605 The sections below describe the logic that was followed in developing the proposed test. As
606 such, many of these sections use general variables (*e.g.* T, L, k, *etc.*). The final numerical
607 values determined for these variables are stated in the beginning of this paper in Section
608 *Proposed Parametric Tolerance Interval Test for Delivered Dose Uniformity Testing of*
609 *OINDP*.

610

611 The fundamental features of the test developed in this section focus on controlling the
612 variability of a single-dose product. The extension to control of potential through-container-
613 life trends for a multi-dose product is dealt with in Section 4, *Sampling Plans*.

614

615 **3.1 Step 1: Introduction of an Acceptance Value**

616 The type of test proposed herein is referred to as a Parametric Tolerance Interval test (PTI test).
617 The primary acceptance criterion for this PTI test is described in terms of an Acceptance Value
618 (AV):

619

$$620 AV = |T - m| + ks$$

621

622 which is required to be not more than a fixed limit (L), *i.e.* $AV \leq L$.

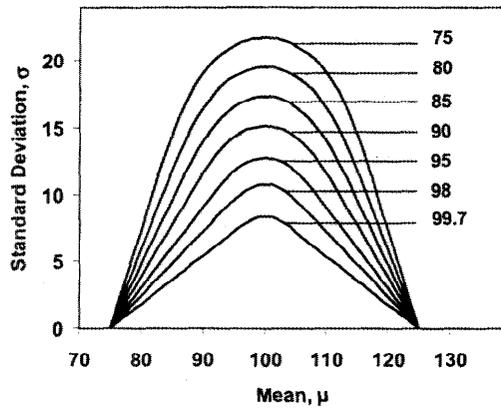
623

624 The Acceptance Value is a linear combination of the absolute deviation of the sample mean
625 (*m*) from the target (*T*) and the sample standard deviation (*s*) scaled by a coefficient (*k*). The
626 limit *L* defines the target interval as $T \pm L$. The scaling coefficient *k* and the sample mean
627 determine the maximum allowable sample standard deviation. The Acceptance Value together
628 with the number of observations (*n*) assesses the quality of the *population* associated with a
629 certain acceptance rate. For convenience, the word *batch* will be used as a synonym for
630 *population*.

631

632 As discussed previously, in the proposed test, quality is defined as the coverage of the target
633 interval, *i.e.* the proportion of the batch that is within the target interval. The coverage of the
634 target interval is fully defined by the mean and standard deviation of a normal distribution and
635 may be calculated by integrating the density distribution between the limits of the target
636 interval¹⁸. The coverage decreases as the mean moves away from the target and/or the standard
637 deviation increases. To maintain constant coverage, therefore, the standard deviation needs to
638 be reduced as the mean moves away from the target, representing a classical trade-off between
639 the mean and variance. This is illustrated in Figure 1, which shows a selection of such *iso-*
640 *coverage* curves for the interval 75-125.

¹⁸ See "Coverage" in the Glossary (Appendix 5) for a formula.



641

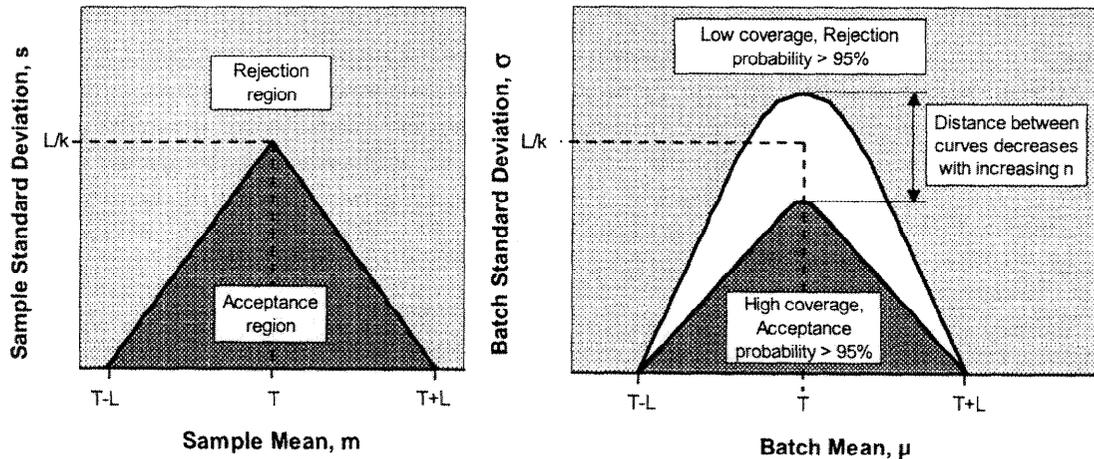
642 **Figure 1. Iso-coverage curves (70-99.7% coverage) for the interval 75-125 for normal**
 643 **distributions.**

644

645 The relation between acceptance rate, coverage, L , T , m and s may be illustrated with reference
 646 to Figure 2. Both panels of Figure 2 show a diagram of Standard Deviation *versus* Mean. The
 647 left panel pertains to sample characteristics, while the right panel pertains to batch
 648 characteristics. It is assumed that the batch is normally distributed with mean μ and standard
 649 deviation σ .

650

651



652

653 **Figure 2. Graphical illustration of the mechanics of a PTI test, see text for details.**

654

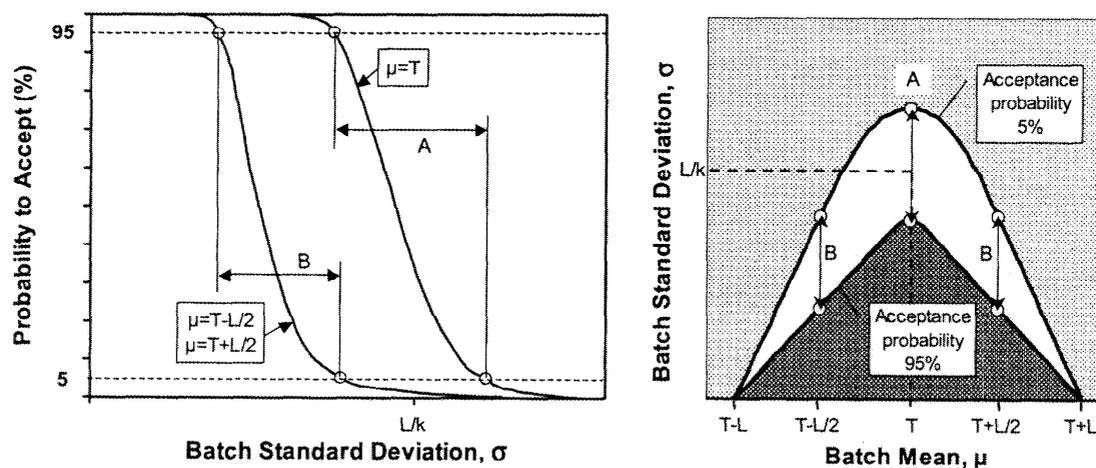
655 Graphically, the combinations of sample standard deviations and sample means that fulfill the
 656 criterion for the Acceptance Value ($AV \leq L$) are delineated by a triangle which defines the
 657 border of the sample acceptance region (Figure 2, left panel). The base of the triangle spans
 658 $T \pm L$, and the height is L/k . A sample with m and s falling within the triangle passes the test,

659 whereas a sample falling outside the triangle fails the test. For an infinitely large sample,
 660 which perfectly reflects the batch characteristics, the triangle would also delineate the quality
 661 of batches accepted or rejected by the test. However, because real samples are necessarily
 662 limited in size, the association between sample and batch characteristics is imperfect (due to
 663 statistical sampling error). For a random sample from a batch within the triangle, there is a
 664 certain probability that the sample characteristics nevertheless will be outside the triangle and
 665 the batch therefore will be falsely rejected (the producer risk). Similarly, for a batch outside
 666 the triangle, there is a probability that the sample characteristics will fall inside the triangle and
 667 the batch therefore will be falsely accepted (the consumer risk).
 668

669 In the right panel of Figure 2, the upper curve shows the combinations (μ , σ) that correspond to
 670 the quality of batches that have exactly 5% probability to provide a sample that passes the test
 671 (such a curve is called an *iso-probability curve*). A batch above this iso-probability curve has
 672 less than 5% chance of passing the test. In other words, this curve represents the quality at 5%
 673 consumer risk, *i.e.* the limiting quality.
 674

675 The lower curve in the right panel shows the combinations (μ , σ) that correspond to the quality
 676 at 95% acceptance probability (5% producer risk). A batch below this iso-probability curve
 677 has more than 95% probability to provide a sample which passes the test.
 678

679 The 5% and 95% acceptance probability curves are derived by calculating the Operating
 680 Characteristic (OC) curves¹⁹ of the test for different batch means, as illustrated in Figure 3. For
 681 each batch mean, the standard deviations giving 5% and 95% acceptance probability (as found
 682 from the OC curve) are plotted on the graph of batch standard deviation *versus* the
 683 corresponding mean. By calculating the OC curve for each of a number of different batch
 684 means, the iso-probability curves corresponding to 5% and 95% acceptance probability can be
 685 constructed.
 686
 687



688

689 **Figure 3. Derivation of the 5% and 95% iso-probability curves from Operating Characteristics**
 690 **curves (see text for details).**

¹⁹ The OC curve used here is a plot of acceptance probability *versus* batch standard deviation for a fixed batch mean.

691
692 The left panel of Figure 3 shows two OC curves; one for a batch with the mean at target ($\mu=T$),
693 and another for a batch with a mean deviating from the target by an amount of $L/2$ ($\mu=T-L/2$),
694 which is identical to the curve for $\mu=T+L/2$. In Figure 3, left panel, the arrowheads of “A”
695 represent the two standard deviations giving 5% and 95% acceptance probability when the
696 mean is at target. The arrowheads of “B” represent the similar quantities when the mean is
697 deviating from the target by $\pm L/2$. In the right panel of Figure 3, these standard deviations are
698 now plotted *versus* the corresponding batch mean.
699

700 3.2 Step 2: Extension to Two-tiered Testing

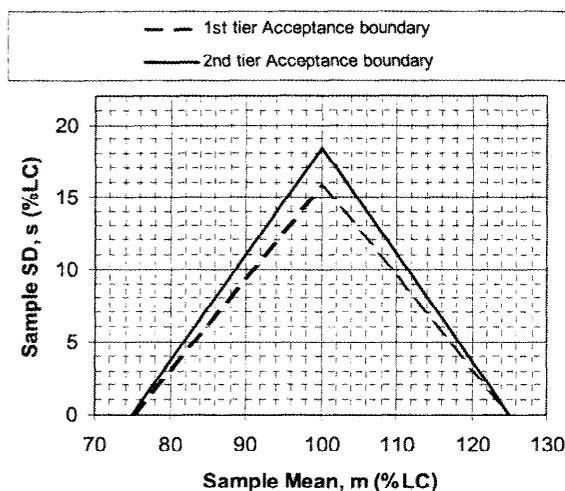
701 To reduce the required number of observations when quality is excellent, a two-tiered test is
702 proposed. The 5% consumer risk for a false acceptance is equally distributed between the two
703 tiers. The distribution of risk is achieved by using different k values for the two tiers. The k
704 value used for the first tier (k_1) is higher, and it restricts the acceptance probability for batches
705 at the limiting quality to 2.5%. If the batch is not accepted in the 1st tier, 2nd tier testing using
706 additional observations is performed. The k value used for the 2nd tier (k_2) is lower, and it
707 allows to accept the remaining 2.5 % of the batches at the limiting quality, for an overall
708 acceptance probability of 5%.
709

710 In the left panel of Figure 2, it can be seen that for a given L , the height of the triangular-
711 shaped acceptance boundary is controlled by the k value. Because k_1 and k_2 are different, the
712 1st and 2nd tier acceptance boundaries differ. Compared to the 1st tier, a larger sample standard
713 deviation is allowed for acceptance in the 2nd tier because the number of observations is higher
714 (which provides a better estimate of batch quality).
715

716 This is illustrated in Figure 4, which shows the 1st and 2nd tier sample acceptance boundaries
717 for a two-tiered PTI test (compare with the left panel of Figure 2). This test comprises 24
718 observations in the first tier (n_1), has a total sample size (n_2) of 72 observations for both tiers,
719 and uses $L=25$, $k_1=1.59$ and $k_2=1.36$.
720

721 Figure 5 shows a number of combinations of batch standard deviation and mean which result in
722 5% (open circles) and 95% (closed squares) acceptance probability for the complete test (*i.e.*,
723 2nd tier testing is employed if a batch was not accepted in the 1st tier).
724

725 Figure 5 also shows two iso-coverage curves for the 75-125% LC target interval: 85% coverage
726 (thick line) and 94.6% coverage (thin line). (The 85% iso-coverage curve represents the
727 limiting quality that has been selected for the proposed test). As is evident from Figure 5, there
728 is a close association between the 5% iso-probability curve and the 85% iso-coverage curve
729 when the batch mean is off target. However, this association is weaker for batch means close
730 to target. In the area near the apex of the curve (where batch means are close to the target), the
731 5 % acceptance probability extends to higher standard deviations than justified by the 85%
732 coverage. Thus, at this step, the acceptance probability for batches close to target is higher
733 than 5% for a coverage of 85%.
734
735



736

Figure 4. Sample acceptance boundaries for 1st and 2nd tier using $L=25$, $k_1=1.59$ and $k_2=1.36$.

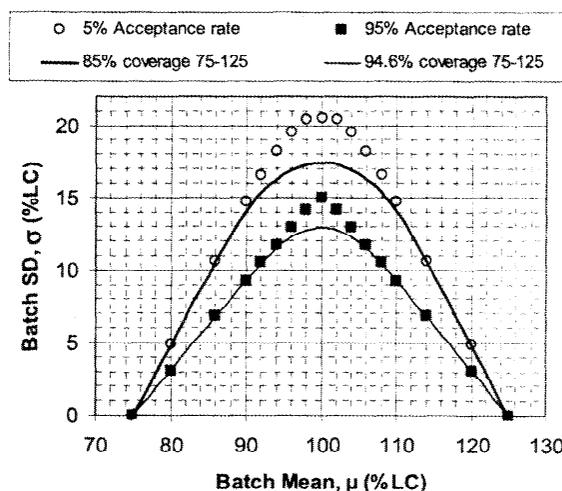


Figure 5. Iso-coverage curves and probability to accept for a two-tiered $n=24/72$ PTI test with acceptance boundaries as per Figure 4.

737

738 This represents a deviation from our goal, namely to design a test that provides close
 739 correlation between acceptance probability and coverage, and which more specifically yields
 740 5% acceptance probability for batches having a coverage of 85%. A PTI test using only the
 741 Acceptance Value as the test criterion does not completely achieve this goal. Therefore, the
 742 PTI test was modified to address this discrepancy, as described in the next section.
 743

744 3.3 Step 3: Introduction of Maximum Sample Standard 745 Deviation

746 The dissociation between the 5% iso-probability curve and the 85% iso-coverage curve in the
 747 vicinity of the target (Figure 5) is due to the fact that the triangular-shaped sample acceptance
 748 boundary (Figure 4) is a simplification of the ideal acceptance criterion. The ideal sample
 749 acceptance criterion has a complex analytical form and can be represented by a triangle with a
 750 rounded, and thus lowered, apex²⁰.

751

752 The complexity of the ideal sample acceptance criterion renders it highly impractical.
 753 Fortunately, the iso-probability curve can be made to trace the iso-coverage curve to a high
 754 degree of accuracy when a criterion that the sample standard deviation may not exceed a
 755 certain maximum is added. This maximum sample standard deviation (MSSD) is conveniently
 756 expressed as a fraction (f) of the height of the triangle:

²⁰ This is due to the fact that for means close to target, individual values may fall outside of the target interval on both sides, which therefore gives a double limitation on the standard deviation, which is not accounted for by the simple triangle. By contrast, for batch means far off target, individual observations that fall outside of the target interval are likely to do so on one side only, and a single limitation on the standard deviation is therefore sufficient.

757

758 $MSSD = fL/k$.

759

760 The sample acceptance boundary created by simultaneously applying criteria on the
761 Acceptance Value and the sample standard deviation can graphically be represented by a
762 truncated triangle, see Figure 6.

763

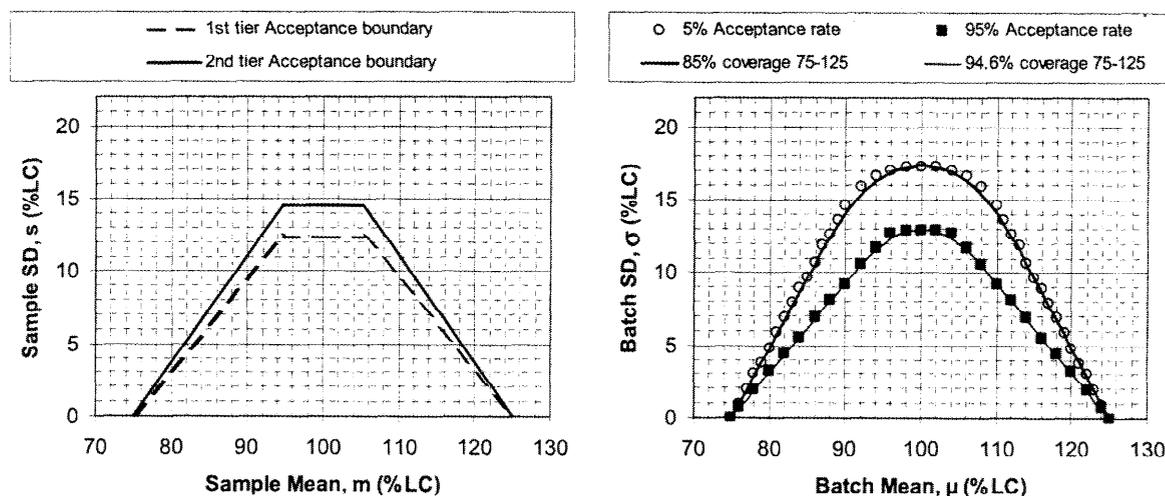
764 The effect on acceptance probability of adding $s \leq MSSD$ as an acceptance criterion is
765 illustrated in Figure 7, which shows the same test as in Figure 5 except for the added MSSD
766 criterion. For this test, using $n=24/72$, the iso-probability curve accurately traces the iso-
767 coverage curve, both at 5% and 95% acceptance rate, when the f value is 0.796 (the value of f
768 varies slightly with sample size).

769

770 Thus, by using the two acceptance criteria (for the sample Acceptance Value and the sample
771 standard deviation), the goal of providing 95% assurance that a batch at the limiting quality
772 will be rejected by the test, has been achieved.

773

774



775

775 Figure 6. Sample acceptance boundaries for 1st and 2nd tier using $L=25$, $k_1=1.59$, $k_2=1.36$ and $f=0.796$.

775 Figure 7. Iso-coverage curves and probability to accept for a two-tiered $n=24/72$ PTI test with acceptance boundaries as per Figure 6.

776

777 3.4 Step 4: Extension to Multiple Test Plans

778 The test proposed here allows the producer to choose from a number of possible test plans,
779 each of which ensures the same limiting quality. The test plans contain a varied number of
780 observations, starting from $n=10/30$. This allows for flexibility in testing the great diversity of
781 orally inhaled and nasal drug products, *i.e.* solution or suspension pMDIs, pre-metered and
782 device metered DPIs, aqueous sprays, *etc.*

783

784 It is the responsibility of the producer to establish the test plan that is most appropriate based
 785 on typical product quality, business needs, and other considerations. An increase in the number
 786 of observations is directly correlated to a decrease in the producer risk and an increase in
 787 required analytical and other resources. A product of excellent uniformity can reap the benefit
 788 of having to test fewer samples without negatively impacting the producer risk. As described
 789 above, the values of k_1 , k_2 and f must be varied with sample size to ensure that there is also no
 790 impact on consumer risk, *i.e.* that a coverage of 85% of the target interval of $100 \pm 25\%$ LC is
 791 associated with an acceptance probability of 5% regardless of the sample size.

792

793 In developing the test plans, the following constraints were used for each sample size:

794

- 795 • The 1st tier acceptance probability is 2.5% for a coverage of 85% of the target interval
 796 $100 \pm 25\%$ LC when the true mean is at 100% LC;
- 797 • The sample size required to be tested in the 2nd tier is twice that in the 1st tier, so that the
 798 total sample size, n_2 , is three times that of the 1st tier ($n_2 = 3n_1$); and
- 799 • The combined acceptance probability for both tiers is 5% for a coverage of 85% of the
 800 target interval $100 \pm 25\%$ LC.

801

802 The acceptance criteria for all test plans are:

803

804 For a sample size of n_1 accept in 1st tier if

805

$$806 \quad |100 - m| + k_1 s \leq 25, \text{ and}$$

807

$$s \leq f 25 / k_1.$$

808

809 If not accepted, proceed to 2nd tier. For a total sample size of n_2 , accept in 2nd tier if

810

$$811 \quad |100 - m| + k_2 s \leq 25, \text{ and}$$

812

$$s \leq f 25 / k_2.$$

813

814 Above, m is the sample average (% LC) and s is the sample standard deviation (% LC).

815

816 Using simulations, the two acceptability coefficients, k_1 and k_2 , and the f factor were
 817 determined for a number of sample sizes using the constraints listed above. The algorithm and
 818 computer code used for these calculations are provided in Appendix 4.

819

820 Table 1 lists the derived coefficients and provides additional information about certain
 821 properties of the test plans. As designed, all test plans have a 5% acceptance probability for a
 822 coverage of 85% of the target interval $100 \pm 25\%$ LC (column *vii*). The required coverage for
 823 95% acceptance probability is much higher, *i.e.* between 94.6% and 97.7% for the six listed test
 824 plans (column *viii*). Note that as coverage decreases (*e.g.* from 97.7% to 94.6%) the producer
 825 risk can be maintained by using a test plan with an increased number of observations. The
 826 mean number of observations at the 95% acceptance probability is slightly more than twice the
 827 size of the 1st tier sample for all test plans (column *xii*).

828

829 The batch standard deviation (σ) corresponding to 5% and 95% acceptance probabilities for a
 830 batch with the mean at target ($\mu = 100\%$ LC) is given in columns *ix* and *x*. For an acceptance
 831 probability of 5% (5% consumer risk) the standard deviation is 17.4% LC for all test plans.
 832 The batch standard deviation for an acceptance probability of 95% (5% producer risk) is much

833 lower, *i.e.* between 11.0% and 12.9% LC. Note again that as the batch standard deviation
 834 increases, the number of observations needs to be increased to maintain the producer risk.
 835

836

837

Table 1. Details and properties of six two-tiered PTI test plans*

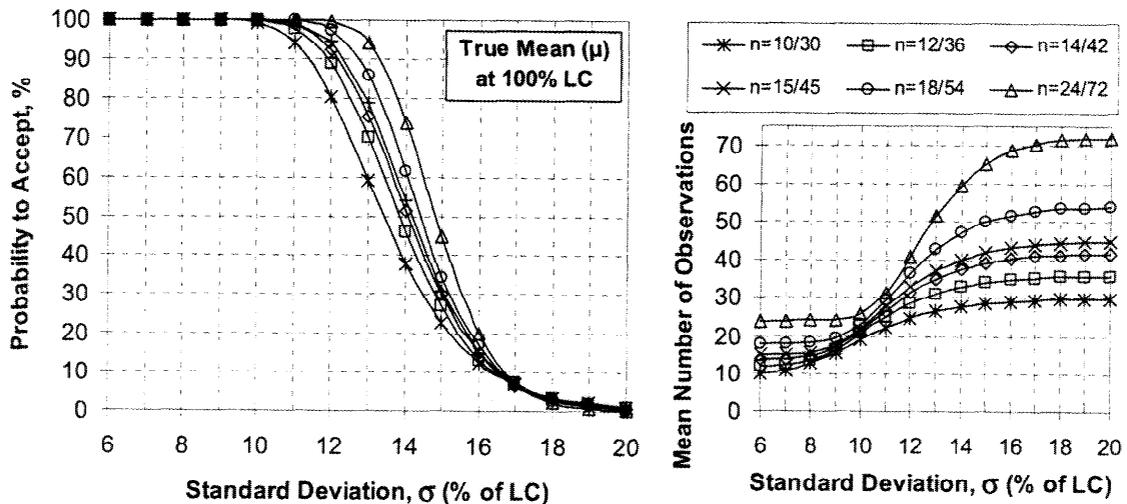
Test plan (i)	n_1 (ii)	n_2 (iii)	k_1 (iv)	k_2 (v)	f (vi)	Coverage (%) of target interval for acceptance probability of		$\sigma_{\mu=100}$ (% LC) for acceptance probability of		Mean number of observations for acceptance probability of	
						5% (vii)	95% (viii)	5% (ix)	95% (x)	5% (xi)	95% (xii)
1	10	30	2.09	1.59	0.839	85	97.7	17.4	11.0	29	22
2	12	36	1.95	1.52	0.826	85	97.0	17.4	11.5	35	26
3	14	42	1.85	1.48	0.819	85	96.7	17.4	11.7	41	30
4	15	45	1.81	1.46	0.815	85	96.4	17.4	11.9	45	33
5	18	54	1.72	1.42	0.808	85	95.6	17.4	12.4	53	38
6	24	72	1.59	1.36	0.796	85	94.6	17.4	12.9	71	51

838

839

840

* Note, these are the same test plans provided in the *Table of Test Plans* (page 7)



841

842 **Figure 8. Left panel: OC curves for the PTI tests described in Table 1. Right panel: Mean**
 843 **number of observations for the tests**

844

845 In Figure 8, the probability to accept a batch (left panel), and the mean number of observations
 846 needed to reach a decision (right panel), are plotted as a function of the batch standard
 847 deviation for six PTI test plans (for batch means at target). The left panel of Figure 8 shows
 848 that for a given batch quality (*here*, standard deviation), the producer risk is progressively
 849 lowered as the sample size is increased (provided, of course, that the batch quality is better than
 850 the limiting quality).
 851

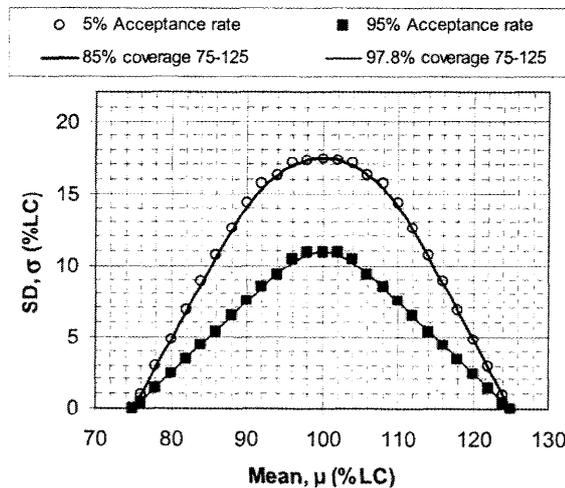
852 The right panel of Figure 8 shows that for each of the PTI test plans, the mean sample size
 853 increases with increasing standard deviation as a consequence of 2nd tier testing becoming

854 progressively more common. For a sufficiently low standard deviation, the test is always
 855 passed in the 1st tier. For a sufficiently high standard deviation, the test always proceeds to the
 856 2nd tier.
 857

858 It is expected that 2nd tier testing will be rather frequent for the PTI test, because the 1st tier
 859 acceptance coefficient, k_1 , has been set to give only 2.5% acceptance probability for the
 860 limiting batch quality (the corresponding acceptance probability for the complete test,
 861 including 2nd tier testing when required, is 5%). The batch is only accepted if the 1st tier
 862 sample meets this more stringent requirement. By contrast, for the FDA tests, 2nd tier testing is
 863 infrequent and plays an insignificant role with regard to batch quality assessment (see
 864 Appendix 1).
 865

866 On average, therefore, the proposed PTI test requires a larger sample size than the FDA tests.
 867 This provides a powerful incentive for producers to improve their product quality since sample
 868 size requirements are lower for products of higher quality.
 869

870 The iso-coverage and iso-probability curves for the smallest and largest tests listed in Table 1
 871 are given in Figure 9 and Figure 10, respectively. These figures show that the iso-probability
 872 curves follow the indicated iso-coverage curves to a high degree of accuracy for all batch
 873 means. This demonstrates the excellent correlation between batch quality (coverage of the
 874 target interval) and sample acceptance probability that is provided by the proposed PTI test.
 875
 876



877

Figure 9. PTI test using $L=25$, $n_1=10$, $n_2=30$, $k_1=2.09$, $k_2=1.59$, $f=0.839$

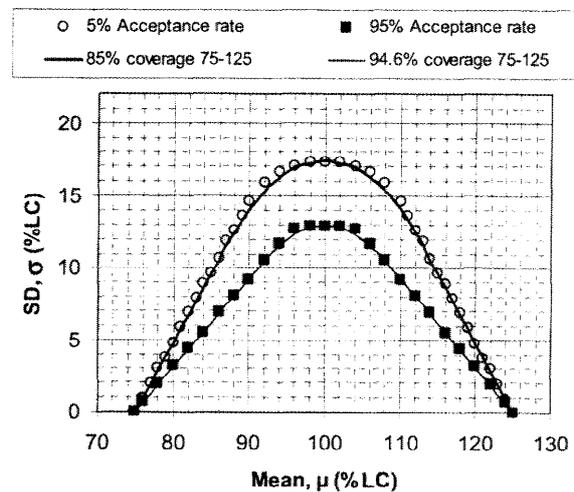


Figure 10. PTI test using $L=25$, $n_1=24$, $n_2=72$, $k_1=1.59$, $k_2=1.36$, $f=0.796$

878

879

880 3.5 Step 5: Introduction of Requirement on Sample 881 Average

882 As is evident from Figure 9 and Figure 10, the PTI test with only two acceptance criteria would
883 accept batches with highly deviating means (up to $\pm 25\%$ deviation from the target) if the
884 standard deviation is sufficiently low to maintain the limiting coverage (85% coverage for 5%
885 acceptance probability). Even though such batches would comply with the coverage criteria,
886 the quality may be regarded as inadequate due to the large mean deviation from the target.
887

888 Therefore the PTI test was further modified to reject batches with highly deviating means. This
889 was accomplished by supplementing the two acceptance criteria introduced above with a third
890 requirement that the sample average (m) be within $100 \pm 15\%$ LC. The proposed limit,
891 $100 \pm 15\%$ LC, is adopted from the draft Guidance tests. Thus, the acceptance criteria for the
892 PTI test at this step appear as follows:
893

894 For a sample size of n_1 accept in the 1st tier if

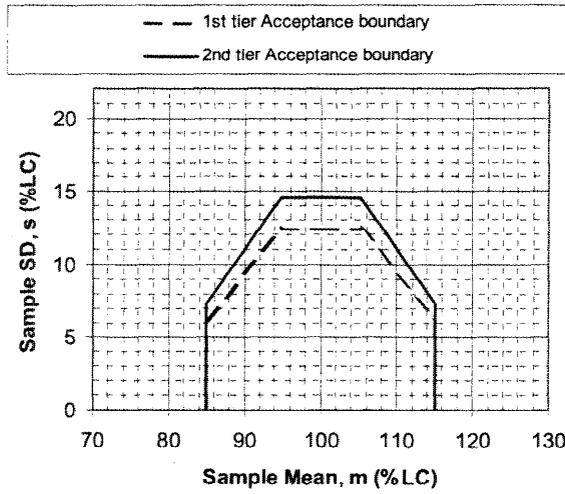
$$\begin{aligned} 895 & |100-m| + k_1 s \leq 25, \\ 896 & s \leq f 25 / k_1, \text{ and} \\ 897 & |100-m| \leq 15. \end{aligned}$$

898
899
900 If not accepted, proceed to the 2nd tier. For a total sample size of n_2 , accept in the 2nd
901 tier if

$$\begin{aligned} 902 & |100-m| + k_2 s \leq 25, \\ 903 & s \leq f 25 / k_2, \text{ and} \\ 904 & |100-m| \leq 15. \end{aligned}$$

905
906
907 The 1st and 2nd tier sample acceptance boundaries created by simultaneously applying the
908 criteria on the Acceptance Value, the sample standard deviation, and the sample average can
909 graphically be represented by triangles truncated at the top and at both flanks, see Figure 11.
910 The effect on acceptance probability of adding the requirement on sample average is illustrated
911 in Figure 12, which shows results for the same test as in Figure 7 except for the added criterion
912 for the sample average. A comparison of these figures clearly shows that batches with large
913 mean deviations and low standard deviations are no longer accepted.
914

915 Note that as with the other acceptance criteria, failure to comply with the criterion for the mean
916 results in a 2nd tier testing. This provides a better estimate of the true mean for the batch and
917 hence is statistically justified.
918



919

Figure 11. Sample acceptance boundaries for 1st and 2nd tier using $L=25$, $k_1=1.59$, $k_2=1.36$, $f=0.796$, and a $100\pm 15\%$ LC requirement on the sample average.

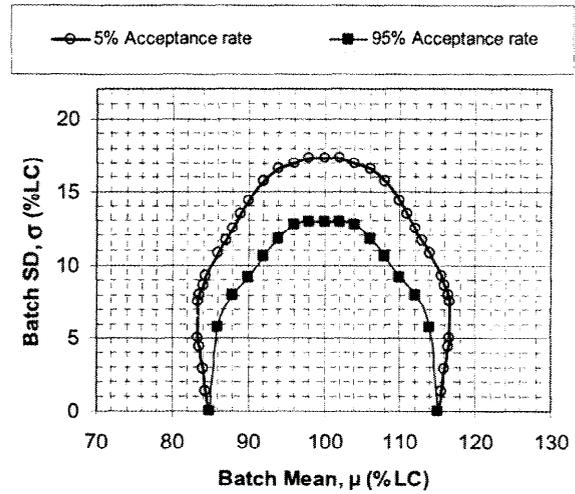


Figure 12. Acceptance probabilities for a two-tiered $n=24/72$ PTI test with acceptance boundaries as per Figure 11

920

921

921

922 **4 Sampling Plans**

923 The sampling plan is an integral part of a test, as it determines what sources of variability will
924 affect the outcome of the test. Since the primary issue at hand is to exercise control over
925 variability of doses within a batch, it is desired to sample doses in such a way that all of the
926 potential sources of variability in the batch are suitably represented by the sample.
927

928 **4.1 Sampling Plan for Single-dose Products**

929 Single-dose products are defined as products in containers that hold a single individually
930 packaged pre-metered dose unit. The delivery mechanism may be an integral part of the
931 container or be provided in an auxiliary device. For a single-dose product, the sampling is
932 straightforward: a representative sample of the containers is tested using the sample size
933 selected for the product. In the case where an auxiliary device is used to deliver the dose from
934 the container, superior control over variability is achieved if a separate delivery device is used
935 to test each dose, because the delivery device may contribute to the overall variability. A
936 separate device should therefore typically be used to test each dose unless it has been
937 demonstrated that an alternative test plan provides equivalent control.
938

939 **4.2 Sampling Plan for Multi-dose Products**

940 Multi-dose products are defined as products in containers that hold multiple doses, whether as
941 reservoirs or as ordered assemblies of individually packaged pre-metered dose units. The
942 delivery mechanism may be an integral part of the container or be provided in an auxiliary
943 device. In the case where an auxiliary device is used to deliver the dose from the container, a
944 separate delivery device should typically be used to test each dose unless it has been
945 demonstrated that an alternative test plan provides equivalent control. The word *inhaler* will
946 be used to denote the container and delivery device combination. For a multi-dose product, it
947 is necessary to sample from different life-stages to evaluate possible systematic trends from the
948 first to the last dose.

949
950 The FDA draft Guidances recommend controlling this potential through-container-life
951 variation by an additional test here denoted as the FDA TCL test²¹, used in addition to the FDA
952 DCU test. In the FDA TCL test, doses from the beginning, middle and end of the container life
953 are sampled from each of three multi-dose inhalers for pMDIs and DPIs (or from the beginning
954 and end from each of five multi-dose inhalers for nasal products and inhalation sprays).
955

956 In the PTI test, it is proposed to control potential through-container-life variation by suitably
957 modifying the PTI sampling plan to a stratified sampling of different life stages for multi-dose
958 products, thus avoiding multiple testing of the uniformity of these products. As described
959 below, the modified sampling plan takes inter- and intra-inhaler variation, including through-
960 life trends, into account simultaneously, and achieves a rigorous overall control of product
961 quality similar to that provided by the combined application of the FDA DCU&TCL tests.
962

²¹ See this entry in the Glossary for an exact reference.

963 4.2.1 Stratified Sampling from Different Life-Stages

964 The objective of a DDU test is to control the overall variability of doses delivered from a batch,
965 irrespective of the sources of variability. The test should be concerned with delivered dose
966 relative to the label claim and not relative to the beginning of a particular inhaler (*i.e.*, for batch
967 control, there is no need to collect beginning, middle, and end doses from the *same* inhaler, as
968 recommended in the draft Guidances). One concept, then, is to view the doses from all life-
969 stages as equally important and test a simple random sample such that each dose at any life-
970 stage and from any inhaler has an equal chance of being part of the sample. In principle, this
971 would ensure the same control over the total variability for multi-dose products as that
972 achieved for single-dose products. However, such a sampling plan would be highly impractical
973 due to logistical difficulties. A more practical way to achieve similar control over the entire
974 population of doses would be to employ a stratified sampling plan where different life-stages
975 provide equal weight.

976
977 The proposed sampling plan for multi-dose products therefore specifies, as a default, that one-
978 third of the doses be sampled from the beginning, one-third from the middle, and one-third
979 from the end of the claimed number of deliveries (inhaler life), each dose being sampled from a
980 unique container and delivery device.

981
982 The requirement that each dose be sampled from a separate inhaler is essential in order to
983 maintain a representative sampling of inhalers. The importance of this requirement depends on
984 the relative magnitude of variation between inhalers and within inhalers. For a case where the
985 between-inhaler variation is much smaller than the within-inhaler variation, the latter would
986 define the overall variability. In this case, the requirement to sample from many different
987 inhalers could be relaxed. However, in order for the proposed test to be generally applicable,
988 the sampling plan was selected so that each dose is sampled from a unique container and device
989 combination.

990
991 After sampling according to the proposed sampling plan, the mean and standard deviation of
992 the total sample (composed of doses from different life-stages) are calculated as described
993 above, and from these the Acceptance Value is derived. The sample standard deviation and
994 Acceptance Value are then compared to the respective acceptance criteria to evaluate the
995 acceptability of the batch. It should be noted that any through-life trend that might be present
996 will inflate the sample standard deviation and thereby also the Acceptance Value. This means
997 that in order to meet the acceptance criteria, the within-life-stage dose-to-dose variation has to
998 be proportionally smaller. This mechanism provides an inherent protection against excessive
999 through-life trends.

1000
1001 The philosophy used here to aggregate different sources of variation and evaluate against a
1002 single metric, in principle follows that of the bioequivalence statistical analysis
1003 recommendations for Dose Content Uniformity Through Container Life²².

1004
1005 In addition to the protection afforded by the inherent sensitivity of the metrics towards life-
1006 stage trends, it is proposed to further limit life-stage trends by requiring that the mean of *each*
1007 life-stage be within 100±15% LC (similar to the requirement of the FDA TCL test) (this
1008 guarantees that the overall mean is also within these limits).

²² *Bioavailability and Bioequivalence Studies for Nasal Aerosols and Nasal Sprays for Local Action*, CDER Draft Guidance for Industry, 1999.

1009

1010 For beginning/middle/end testing, sampling plans with n's divisible by three (e.g., 12/36,
1011 15/45, 18/54, 24/72) are suitable. *Beginning* is defined as the first dose after preparation of the
1012 inhaler according to the label (e.g., after priming maneuvers when so directed), *middle* is the
1013 next dose after delivery of half of the claimed number of doses, and *end* is the last of the
1014 claimed number of doses.

1015

1016 For a product which has been demonstrated to have no trend or a monotonic trend through
1017 container life, it is proposed that testing of the middle life stage be waived. A product is
1018 monotonic if the level of the middle life stage is typically contained within the interval
1019 determined by the levels of the beginning and end life stages. In such cases, one-half of the
1020 doses may be sampled from the beginning and one-half from the end of the container life. The
1021 justification for this is two-fold. Firstly, the sampling plan becomes logistically simpler,
1022 thereby saving analytical resources. Secondly, by testing only the beginning and end of
1023 monotonic products, the test becomes more stringent because the sampling focuses on worst-
1024 case scenarios and avoids diluting of the sample information with doses from the middle life
1025 stage, when this has been shown to be bracketed by the beginning and end doses.

1026

1027 For beginning/end testing, sampling plans with even n's are suitable (e.g., 10/30, 12/36, 14/42,
1028 18/54, 24/72).

1029

1030

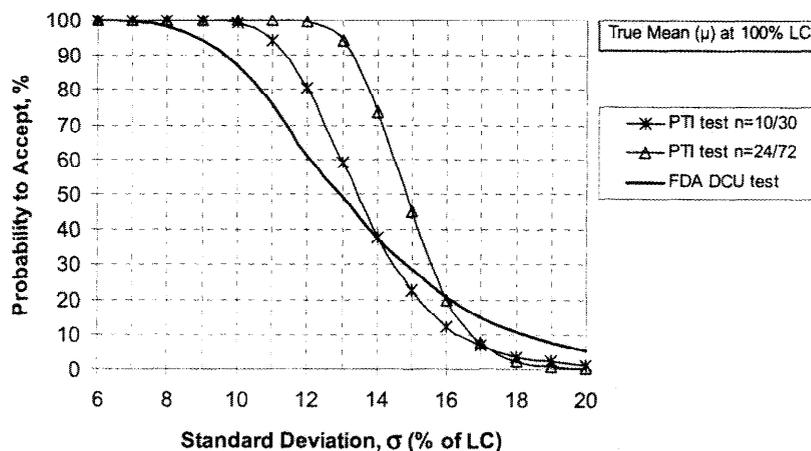
1031

1031

1032 **5 Comparison of PTI Tests with FDA Tests**1033 **5.1 Single-dose Products**

1034 The operating characteristics of the PTI tests with the small ($n=10/30$) and large ($n=24/72$)
 1035 sample sizes described in the *Table of Test Plans* (page 7) were investigated by simulation
 1036 using the acceptance criteria described above. The OC curves for batch means at target are
 1037 given in Figure 13, which as a comparison also shows the OC curve for the FDA DCU test.
 1038 Figure 13 demonstrates that the OC curves for the PTI test (both test plans) are sharper and
 1039 provide both improved consumer and producer protection compared to the FDA DCU test.
 1040

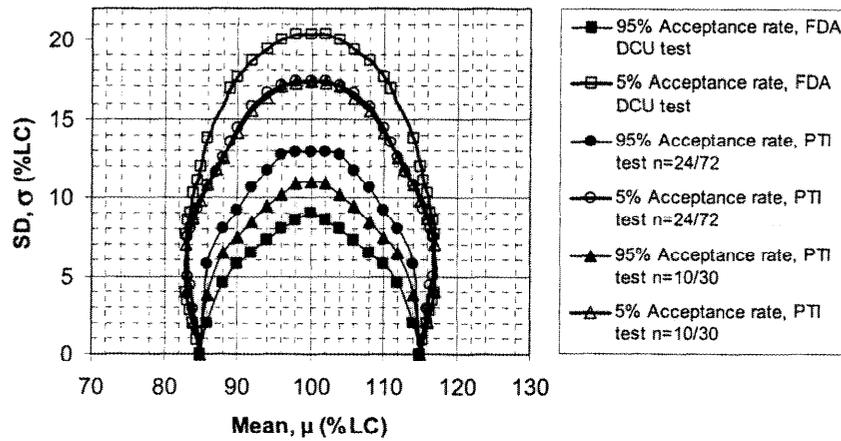
1041 The batch characteristics giving 5% and 95% acceptance probabilities using these tests are
 1042 given in Figure 14. This figure demonstrates that the PTI tests provide better consumer
 1043 protection than the FDA DCU test for all combinations of batch mean and batch standard
 1044 deviation. This is evidenced by the fact that the iso-probability curves for 5% acceptance for
 1045 the PTI test are completely inscribed within the corresponding curve for the FDA DCU test.
 1046



1047

1048 **Figure 13. Comparison of the OC curve for two of the PTI tests described in the *Table of Test***
 1049 ***Plans* (including the $100\pm 15\%$ LC requirement on the sample average), and the FDA**
 1050 **DCU test**

1051



1052

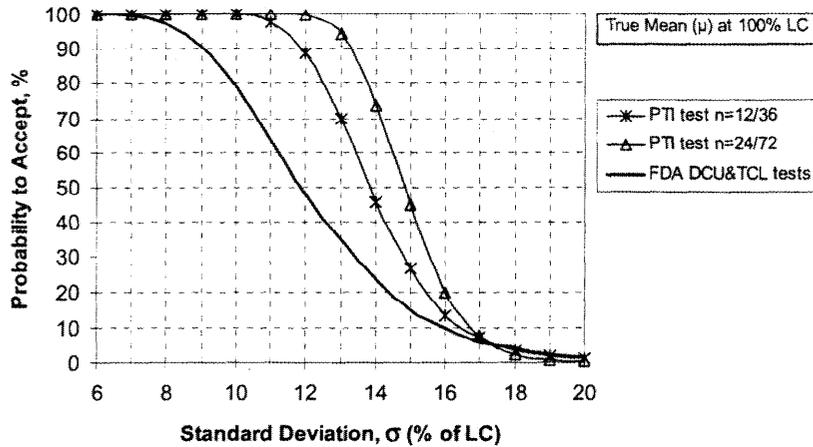
1053 **Figure 14. Comparison of the operating characteristics for two of the PTI tests described in the**
 1054 **Table of Test Plans (including the $100 \pm 15\%$ LC requirement on the sample average),**
 1055 **and the FDA DCU test**

1056

1057 5.2 Multi-dose Products

1058 The operating characteristics of the tests with the small ($n=12/36$, *i.e.*, the smallest test divisible
 1059 by 3) and large ($n=24/72$) sample sizes described in the *Table of Test Plans* (page 7) were
 1060 investigated by simulation using the acceptance criteria described above (including the
 1061 requirement on the mean applied to each life-stage separately). The OC curves for batch means
 1062 at target are given in Figure 15, which as a comparison also shows the OC curve for the FDA
 1063 DCU&TCL test. Figure 15 demonstrates that the OC curves for the PTI test (both test plans)
 1064 are sharper and provide similar consumer protection and improved producer protection
 1065 compared to the FDA DCU&TCL test. The batch characteristics giving 5% and 95%
 1066 acceptance probabilities using these PTI tests are given in Figure 16. This figure demonstrates
 1067 that the PTI tests provide comparable consumer protection to that of the FDA DCU&TCL tests
 1068 for all combinations of batch mean and batch standard deviation. This is evidenced by the fact
 1069 that the iso-probability curves for 5% acceptance for the PTI tests trace the corresponding
 1070 curve for the FDA DCU&TCL test.

1071



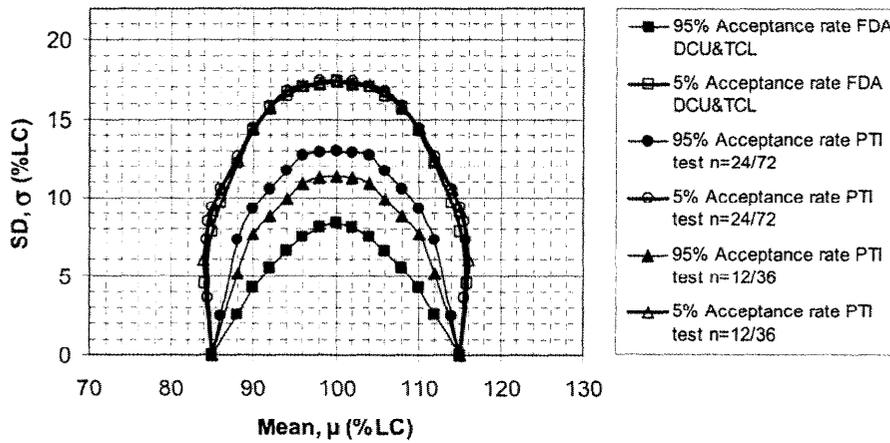
1072

1073

1074

1075

Figure 15. Comparison of OC curve for two of the PTI tests described in the *Table of Test Plans* (including the $100 \pm 15\%$ LC requirement on the sample average for each life-stage separately), and the FDA DCU&TCL tests



1076

1077

1078

1079

Figure 16. Comparison of the operating characteristics for two of the PTI tests described in the *Table of Test Plans* (including the $100 \pm 15\%$ LC requirement on the sample average for each life-stage separately), and the FDA DCU&TCL tests

1080

1081

1081

1082 **6 Sufficiency of PTI Test Criteria**

1083 The previous sections discussed the considerations used to develop the proposed parametric
1084 tolerance interval test, which includes criteria for: 1) the Acceptance Value, 2) the sample
1085 standard deviation, and 3) the sample average (applied to each life-stage separately for multi-
1086 dose products). This section presents further considerations demonstrating that these three
1087 criteria are sufficient to achieve efficient and rigorous control over dose uniformity.
1088

1089 **6.1 The Parametric Criteria of the PTI Test Replaces the**
1090 **Zero-tolerance Criterion of the FDA DCU Test**

1091 The FDA draft Guidance test includes a so-called “zero-tolerance” requirement, *i.e.*, the
1092 requirement that no observed value in a sample may be outside $100\pm 25\%$ LC. A zero-tolerance
1093 requirement, however, does not provide a safety net against the presence of outlying doses
1094 within the batch. It only provides a safety net against outlying doses within the sample. A
1095 batch may still contain a significant percentage of such doses even when the sample does not.
1096 The reason for this is that, statistically, there will always be a risk that the sample is free from
1097 outlying results (by pure chance) even when such doses exist in the batch. The zero-tolerance
1098 requirement constitutes a simple and correct non-parametric tolerance interval test of a
1099 $100\pm 25\%$ LC target interval. However, such a test has a rather flat operating characteristic
1100 curve, which means that the chance to detect (and reject) a batch with an elevated frequency of
1101 outlying doses increases rather slowly with the frequency of outlying doses. The proposed
1102 parametric test has the capability to be more efficient in this respect.
1103

1104 Appendix 1 demonstrates that a batch will meet the FDA zero tolerance requirement with 5%
1105 probability if the coverage is 74%²³. This means that up to 26% of the values in a batch may
1106 be outside of the interval before the probability that at least one is present in the sample ($n=10$)
1107 reaches 95%. At the same time, due to the flatness of the OC curve for this criterion, a batch
1108 containing as little as 0.5% of the values outside the interval, has a 5% risk of being rejected
1109 due to this criterion. Thus, this criterion carries a high risk for the producer without providing
1110 the consumer a high protection.
1111

1112 With the parametric criteria of the PTI test, on the other hand, each accepted batch contains less
1113 than 15% values outside the $100\pm 25\%$ LC interval with 95% confidence. Thus, the parametric
1114 criteria of the PTI test afford superior consumer protection against doses outside the target
1115 interval compared to the zero-tolerance criterion of the FDA DCU test.
1116

1117 The combined application of the FDA DCU and TCL tests for a multi-dose product also
1118 achieves 95% confidence that an accepted batch contains less than 15% values outside of this
1119 interval. Thus, the same consumer protection is achieved with the FDA DCU&TCL test as
1120 with the PTI test. However, the producer risk that a uniform batch is rejected is much higher
1121 with the FDA DCU&TCL tests than with the PTI test.
1122

²³ The 78% coverage quoted earlier results from the application of *all three* FDA criteria (*i.e.*, zero tolerance, $\pm 20\%$ limit, and sample mean).

1123 In Appendix 2 it is demonstrated that overall, real data from orally inhaled and nasal drug
1124 products follow a normal distribution with excellent fit extending far out into the tails of the
1125 distribution. Thus, chance observations located in the extreme tails of the distribution are
1126 expected to occur even for batches with excellent uniformity. This signifies that there exists no
1127 useful limit beyond which a *single* observation in a sample would constitute evidence that a
1128 batch is unacceptable. Institution of a requirement that depends on chance rather than on batch
1129 quality is not ideal for quality control. In this respect, parametric requirements are superior to
1130 non-parametric ones because they focus on overall batch quality rather than on individual
1131 sample observations.
1132

1133 **6.2 The Addition of a Zero Tolerance Criterion to the PTI** 1134 **Test would be Incongruent with the Parametric** 1135 **Approach**

1136 The limiting quality in the proposed approach is defined as an 85% coverage of the $100\pm 25\%$
1137 LC target interval. This implies that doses outside of the target interval are not disallowed.
1138 Because doses outside the target interval are tolerated (at a low, controlled frequency), it would
1139 be inappropriate to add a requirement that a single observation, in a sample, of a dose outside
1140 of the target interval is unacceptable.

1141
1142 The addition of a zero tolerance criterion to the PTI test would degrade the test. Figure 17 and
1143 Figure 18 show the effect on the OC curve of adding a zero-tolerance criterion for the target
1144 interval $100\pm 25\%$ LC for two of the PTI tests in the *Table of Test Plans* (page 7) (those using
1145 the lowest and highest numbers of observations, respectively). The graphs clearly show that
1146 the OC curve for the test including the zero-tolerance requirement is less steep than the OC
1147 curve for the test using parametric requirements only. For large standard deviations, there is
1148 only a slight effect on the acceptance rate whereas for low to moderate standard deviations, the
1149 acceptance rate is dramatically affected by the zero-tolerance requirement, particularly for the
1150 larger sample size. In fact, the OC curve for the large sample is completely defined by the
1151 zero-tolerance requirement, meaning that the parametric criteria no longer affect the shape or
1152 location of the OC curve.

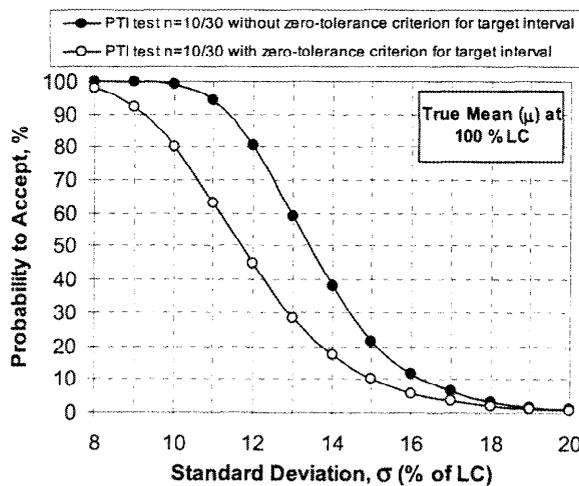
1153
1154 Ideally, as the sample size increases (and thus information about batch quality becomes more
1155 complete), a proper test should increase the acceptance rate for uniform batches, and decrease
1156 the acceptance rate for batches of low uniformity. Contrary to this goal, a test with a zero-
1157 tolerance criterion has a lower acceptance rate for larger samples compared to smaller samples,
1158 as illustrated by comparison of the OC curves in Figure 17 and Figure 18. Here, the OC curve
1159 for the $n=24/72$ sample lies lower than the OC curve for the $n=10/30$ sample for *all* standard
1160 deviations. As described further in Appendix 1, a zero-tolerance requirement will always, for
1161 *any* given quality, decrease the acceptance rate as the number of observations increases. With
1162 a zero-tolerance requirement, every observation carries a random risk to fail, encouraging
1163 minimalistic testing. This has the undesirable consequence that investigations requiring
1164 repeated testing, such as development, stability and validation, are at risk of being under-tested.
1165 Thus, the addition of a zero-tolerance requirement to a parametric test would be counter to the
1166 intent of quality control, as it discourages adequate testing of a batch and thorough assessment
1167 of its true quality.
1168

1169 The addition of a zero-tolerance criterion to the PTI test would compromise two of the key
 1170 elements of the proposed test, namely:

- 1171
- 1172 • the higher efficiency (steeper OC curve, *i.e.*, better discriminatory power) compared to
- 1173 non-parametric tests, and
- 1174 • the ability to mitigate producer risks by increasing the sample size.
- 1175

1176 At the same time, the addition of a zero-tolerance criterion would not meaningfully improve
 1177 the consumer protection, as the level of consumer protection against acceptance of low quality
 1178 batches provided by the parametric criteria of the proposed PTI tests already is superior to that
 1179 given by the FDA draft Guidance DCU test for a single-dose product (see Figure 13 above) and
 1180 comparable to that given by the FDA DCU&TCL tests for a multi-dose product (see Figure 15
 1181 above).

1182 Therefore, it is concluded that the addition of a zero-tolerance criterion to a PTI test is
 1183 unwarranted and highly undesirable as it would provide no added value and is associated with
 1184 many drawbacks. The elimination of a zero tolerance criterion has been fully compensated for
 1185 by the use of the more efficient parametric criteria contained in the PTI tests.
 1186
 1187
 1188



1189

Figure 17. PTI test using $L=25$, $n_1=10$, $n_2=30$, $k_1=2.09$, $k_2=1.59$, $f=0.839$, with or without a zero-tolerance criterion as indicated

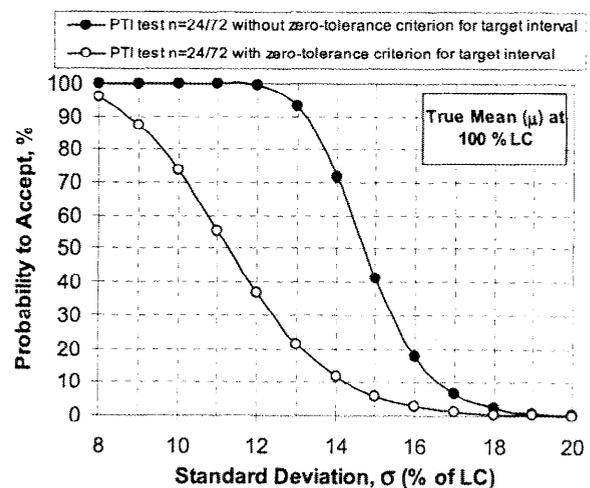


Figure 18. PTI test using $L=25$, $n_1=24$, $n_2=72$, $k_1=1.59$, $k_2=1.36$, $f=0.796$, with or without a zero-tolerance criterion as indicated

1190
 1191

1192 6.3 Non-normal Distributions are Treated Conservatively 1193 by PTI Tests

1194
 1195 The theoretical foundation for the proposed PTI tests is based on the assumption that doses in a
 1196 batch are normally distributed. It is thus important to investigate how non-normal distributions

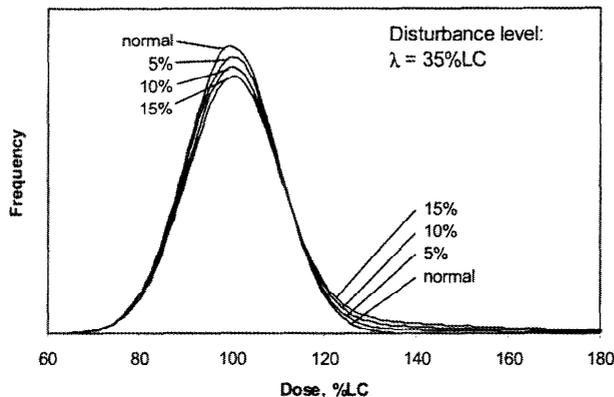
1197 are treated by the PTI test, even though there are strong indications that overall, orally inhaled
1198 and nasal drug products are normally distributed (see Appendix 2).
1199

1200 An interesting type of non-normality arises when the basic, normal distribution is contaminated
1201 with values from another distribution. Here, the effect of such contaminating values will be
1202 studied by disturbing a proportion of the values of a normal distribution by another process. To
1203 investigate whether the parametric criteria alone offer protection against disturbances, a
1204 selection of PTI tests for single-dose products were compared with the FDA DCU test, and a
1205 selection of PTI tests for multi-dose products were compared with the FDA DCU&TCL tests.
1206 The draft Guidance tests were used as references, because they contain a zero-tolerance
1207 criterion for values outside of $100 \pm 25\%$ LC. To facilitate comparison between the tests, which
1208 all have different OC curves, the standard deviation of the basic distribution (a normal
1209 distribution with the true mean at target) was adjusted for each test to give 95% acceptance rate
1210 in the absence of disturbances.
1211

1212 For the purpose of this exercise, it was assumed that disturbances are positive and that a small
1213 disturbance is more common than a larger one. A simple one-parameter distribution that
1214 fulfills these conditions is the exponential distribution (with density function $\exp(-x/\lambda)/\lambda$, $x \geq 0$,
1215 $\lambda > 0$). The parameter λ of the distribution equals both the mean and the standard deviation of
1216 the exponential distribution. For example, for $\lambda = 35\%$ LC, the average size of the disturbances
1217 is $+35\%$ LC. However, because a small disturbance is more common than a larger one, the
1218 median disturbance size is lower; in this example, the median is $+24\%$ LC ($35 \ln(2)$).
1219

1220 The simulations were performed as follows. A value was randomly drawn from the basic,
1221 normal distribution. With a certain probability, this value was then disturbed by the *addition* of
1222 a randomly drawn value from the selected exponential distribution. This was repeated until a
1223 sample size sufficient for final evaluation was reached. The whole procedure was then
1224 repeated several thousand times to obtain sufficient accuracy to draw smooth curves.
1225

1226 Figure 19 compares the density function of an undisturbed normal distribution with those of
1227 normal distributions contaminated with disturbances ($\lambda = 35\%$ LC) with increasing frequency
1228 ($p = 5, 10, 15\%$). As is evident from the figure, the exponentially disturbed distributions have a
1229 heavy right-hand tail and a lower frequency of values near the target.
1230



1231

1232

1233

Figure 19. Density function for a pure normal distribution, and a normal distribution disturbed by an exponential distribution ($\lambda = 35\% \text{ LC}$, probability of a disturbance as indicated).

1234

1235

1236

1237

1238

In one set of simulations, the probability of disturbing a value of the normal distribution was varied (5, 10, 15%) for a fixed level of disturbance ($\lambda = 35\% \text{ LC}$); and in another set, the level of disturbance (λ) was varied (0-100% LC) for a fixed probability (10%).

1239

1240

1241

1242

1243

1244

1245

1246

1247

Figure 20 (left panel: level of disturbance fixed at $\lambda = 35\% \text{ LC}$; right panel: probability fixed at 10%) indicates that both the small and large PTI tests for single-dose products (using no zero-tolerance criterion) react more sensitively to disturbances than the FDA DCU test (which uses a zero-tolerance criterion). For all tests, the sensitivity to disturbances increases with increasing frequency and/or level. The predominant reason why the PTI tests are discriminatory against disturbances in spite of not using a zero-tolerance criterion is the fact that the standard deviation is inflated by the presence of deviating values to make it transgress the MSSD criterion and/or make the Acceptance Value exceed its limit.

1248

1249

1250

1251

1252

1253

1254

Figure 21 shows the results for multi-dose products. The same conclusions as for single-dose products hold, although the degree of improved control achieved by the PTI tests is less pronounced. The reason is that compared to the FDA DCU test for single-dose products, the FDA DCU&TCL combined test for multi-dose products uses a larger sample size and thus naturally has a greater sensitivity towards disturbances than the FDA DCU test. The PTI tests behave similarly for both product types.

1255

1256

1257

1258

1259

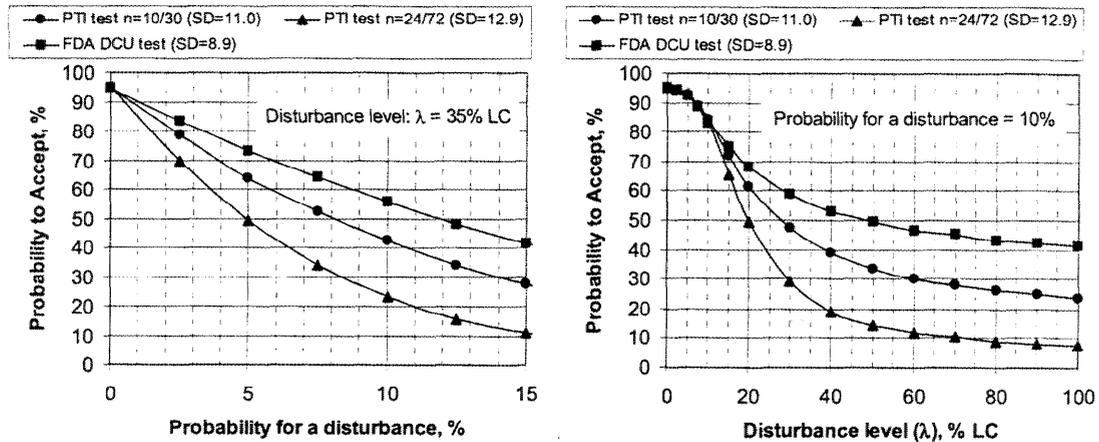
In Appendix 3, it is further demonstrated that the PTI tests are conservative and that for a fixed coverage, the probability of acceptance decreases when data are non-normally distributed. This is shown for the main classes of non-normal distributions potentially encountered in practice (skewed distributions, multi-modal distributions and heavy-tailed distributions).

1260

1261

1262

These findings show that consumer protection is not eroded when the PTI test is applied to non-normal data. Thus, it is valid to use the PTI test both for normal and non-normal data.



1263

1264

1265

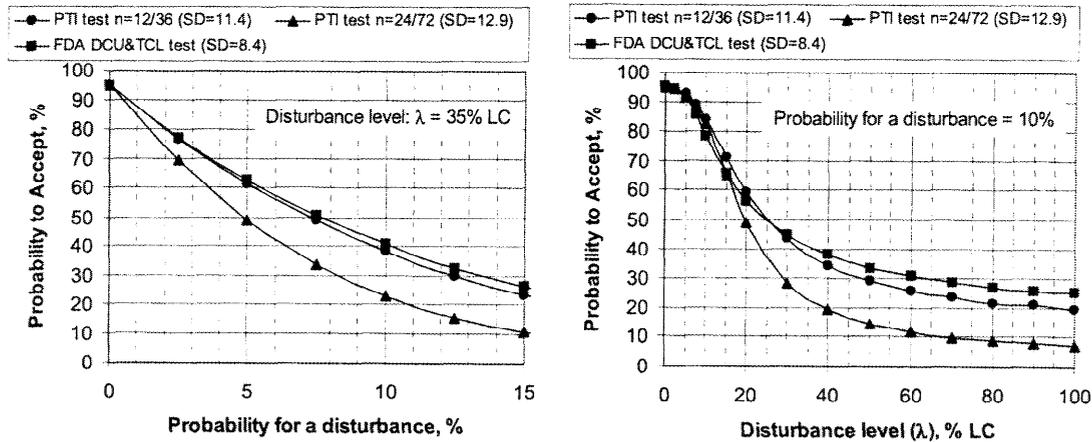
1266

1267

Figure 20. Acceptance probability for single-dose products as a function of disturbance probability (left panel) and disturbance level (λ) (right panel). Basic (normal) distribution with mean at target and standard deviation adjusted (see legend) to give 95% acceptance probability in absence of disturbances. See text for details.

1268

1269



1270

1271

1272

1273

1274

Figure 21. Acceptance probability for multi-dose products as a function of disturbance probability (left panel) and disturbance level (λ) (right panel). Basic (normal) distribution with mean at target and standard deviation adjusted (see legend) to give 95% acceptance probability in absence of disturbances. See text for details.

1275

1276

1276

1277 7 Simulated Production Situation

1278 Although formally described by the operating characteristic curves, the practical capability of a
1279 test may not be immediately apparent. An illustrative way to study the capability of a test is to
1280 try to simulate a realistic production situation, where a series of batches of varying quality is
1281 inspected, and then separately study the quality of batches that were rejected and accepted by
1282 the test under consideration. It should be emphasized that it is not the *quantity* of rejected
1283 batches, but rather the *quality* of *accepted* batches that should be the focus of proper quality
1284 control.

1285

1286 7.1 Simulated Single-dose Product

1287 Figure 22 shows an example for a simulated production of a single-dose product yielding
1288 batches with a long-term (*i.e.*, over many batches) average of 100% LC and a long-term within-
1289 batch standard deviation of 10% LC.

1290

1291 To simulate a production where the true quality of batches varies, the following procedure was
1292 used. For each simulated batch (batch number i), a true mean (μ_i) was randomly drawn from a
1293 normal distribution with the mean at 100% LC and a standard deviation of 4.5% LC (*i.e.*
1294 $N(100, 4.5)$). In other words, the overall product mean (over all batches) is 100% LC and the
1295 standard deviation of batch means is 4.5% LC. Similarly, for each batch the true within-batch
1296 standard deviation (σ_i) was randomly drawn from a normal distribution $N(10, 1.5)$. That is, the
1297 overall within-batch standard deviation (over all batches) is 10% LC and the variability of the
1298 within-batch standard deviation is 1.5% LC. The figures of 4.5%, 10% and 1.5% are arbitrary
1299 but considered to be realistic.

1300

1301 Thus, by this procedure, each simulated batch has a known true mean (μ_i) and a known true
1302 standard deviation (σ_i). Further, assuming that values within a batch follow the normal
1303 distribution $N(\mu_i, \sigma_i)$, values can then be randomly drawn from this distribution to simulate a
1304 sample from batch i . Finally, an FDA or a PTI test can be applied to these values.

1305

1306 The above procedure was applied to 5000 simulated batches. Figure 22 (page 43) illustrates
1307 the true properties (μ_i and σ_i) of batches that passed or failed the FDA DCU test and two of the
1308 PTI tests ($n=10/30$ and $n=24/72$) based on results from a random sample from each batch.

1309

1310 Inspection of the true quality of the batches accepted by the tests shows that there was no
1311 meaningful difference in the true quality of batches accepted by the FDA and PTI tests. The
1312 median coverage of the 75-125% target interval of accepted batches was 98.4% (5 to 95
1313 percentiles: 94.6-99.9% coverage) for the FDA DCU test, 98.3% (5 to 95 percentiles: 94.7-
1314 99.9% coverage) for the small PTI test, and 98.1% (5 to 95 percentiles: 93.9-99.8% coverage)
1315 for the large PTI test. For this simulated situation, the FDA DCU test rejected about 25% of
1316 the batches. The PTI tests rejected 13% (small sample) and 4% (large sample). The FDA
1317 DCU test thus rejected a significantly higher number of batches than the PTI tests, yet the
1318 outgoing batch quality was the same.

1319

1320 The quality of *rejected* batches, on the other hand, was clearly different. As is illustrated in
1321 Figure 22, the FDA DCU test rejected a significant fraction of batches that are well within the

1322 region of quality *accepted* by the FDA DCU test. In contrast, the PTI tests exhibited a better
1323 discrimination between the quality of accepted and rejected batches (of course, this is due to
1324 the fact that the FDA DCU test has a shallower OC curve than the PTI test, see Figure 13).
1325

1326 The higher discriminating power of the PTI test comes at the expense of more testing. For this
1327 particular simulation, the FDA DCU test used on average 11.1 observations per test to assess a
1328 batch. By contrast, the PTI test used on average 20.4 (small test) or 32.7 (large test)
1329 observations per test, due to the 2nd tier testing being invoked for 52% of the batches using the
1330 small PTI test, and 18% using the large PTI test.
1331

1332 7.2 Simulated Multi-dose Product

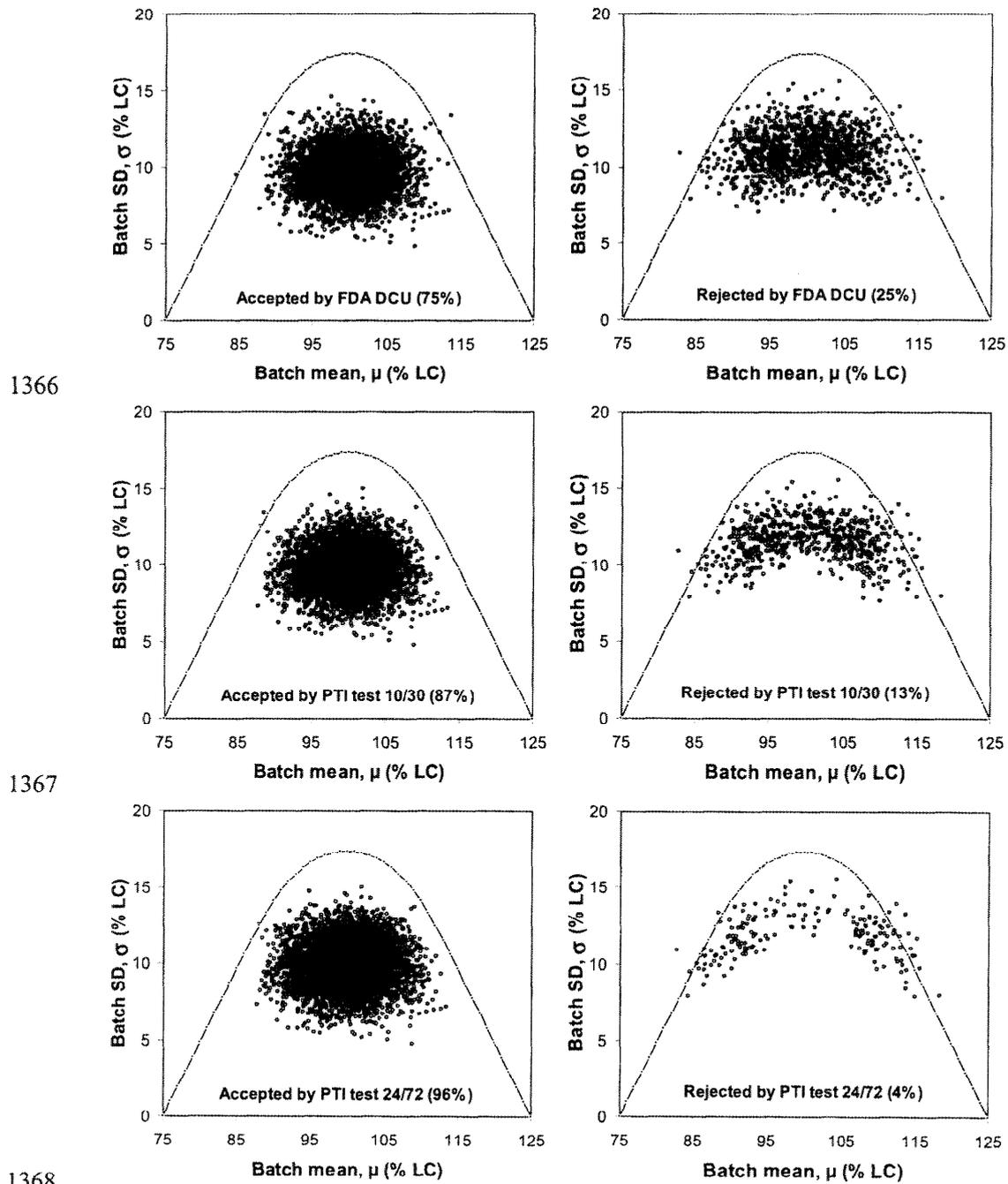
1333 Figure 23 shows an example of a simulated production of a multi-dose product, also yielding
1334 batches with a long-term (over many batches) average of 100% LC and a long-term within-
1335 batch standard deviation of 10% LC, again assuming the variability between batch means and
1336 standard deviations to be 4.5% and 1.5%, respectively. For this example, it was assumed that
1337 the within-batch variability originated in equal parts from intra- and inter-container variability
1338 and that there was no systematic through-container-life trend.
1339

1340 Figure 23 illustrates the true properties (μ_i and σ_i) of batches that passed or failed the FDA
1341 DCU&TCL combined test and two of the PTI tests ($n=12/36$ and $n=24/72$, using equal
1342 sampling in the beginning, middle and end of container life) as judged from a random sample
1343 from each batch.
1344

1345 An inspection of the true quality of the batches accepted by the tests again shows that there was
1346 no appreciable difference in quality of batches accepted by the FDA and PTI tests. The median
1347 coverage of the 75-125% target interval of accepted batches was 98.6% (5 to 95 percentiles:
1348 95.1-99.9% coverage) for the FDA DCU&TCL test, 98.3% (5 to 95 percentiles: 94.5-99.9%
1349 coverage) for the small PTI test, and 98.1% (5 to 95 percentiles: 93.9-99.8% coverage) for the
1350 large PTI test. The FDA DCU&TCL test rejected about 35% of the batches. The PTI tests
1351 rejected 10% (small sample) and 4% (large sample) of the batches. Thus, the FDA DCU&TCL
1352 test rejected a significantly higher number of batches than the PTI tests, with the same outgoing
1353 quality.
1354

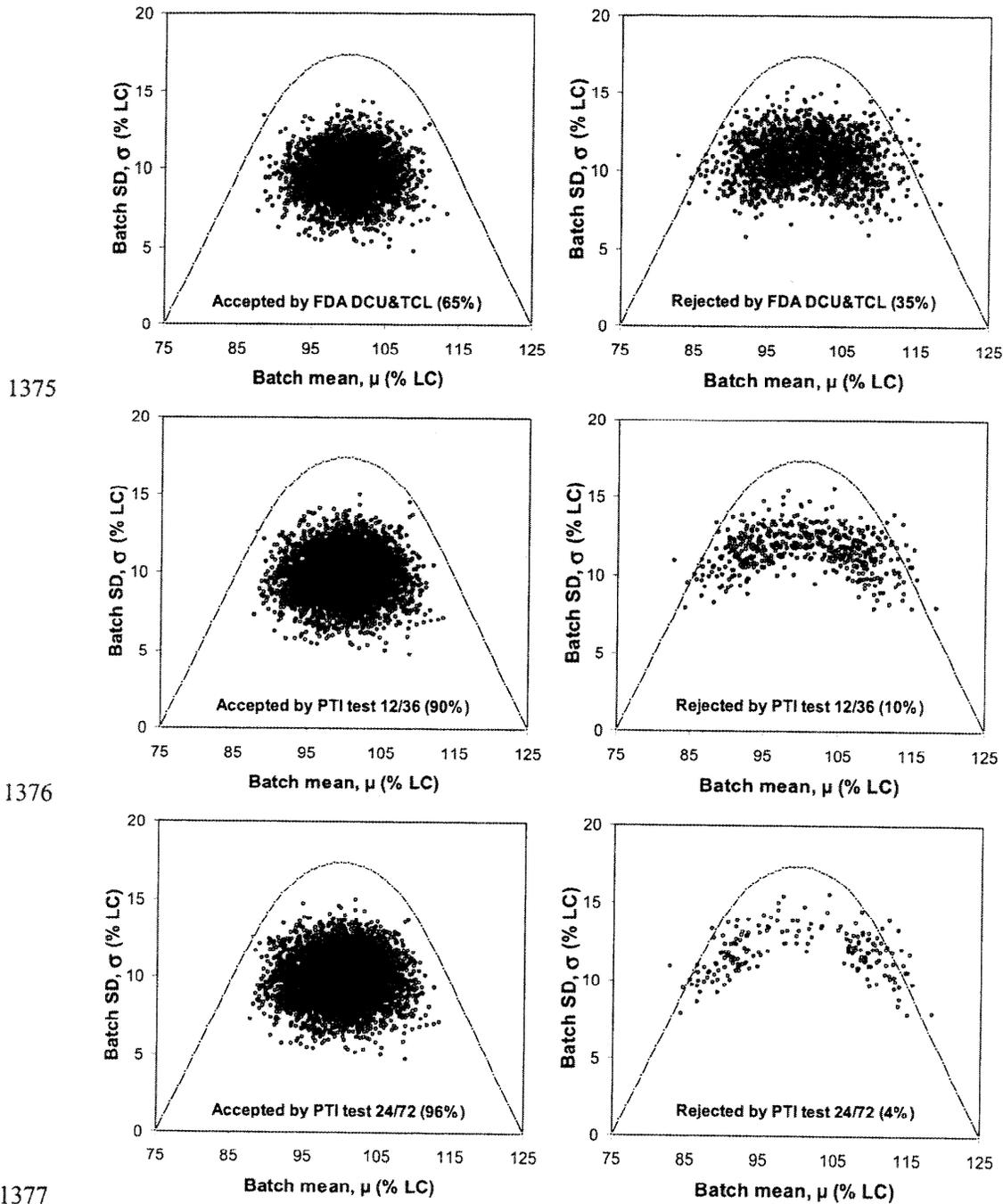
1355 Again, the quality of the rejected batches was clearly different. As is illustrated in Figure 23,
1356 the FDA DCU&TCL test rejected a significant fraction of batches that are well within the
1357 region of quality accepted by the FDA DCU&TCL test. In contrast, the PTI tests exhibited a
1358 better discrimination between the quality of accepted and rejected batches. (Again, this is due
1359 to the fact that the FDA DCU&TCL test has a shallower OC curve than the PTI test, see Figure
1360 15).
1361

1362 The FDA DCU&TCL test used on average 17.7 observations per test to assess the batch
1363 quality. By contrast, the PTI test used on average 22.9 (small test) or 32.8 (large test) values
1364 per test, due to 2nd tier testing being invoked for 45% of the batches using the small PTI test
1365 and 18% using the large PTI test.



1369 **Figure 22.** Comparison of two of the PTI tests described in the *Table of Test Plans* (including
 1370 the 100±15% LC requirement on the sample average), and the FDA DCU test in a
 1371 simulated production situation for a single-dose product. Each dot represents the
 1372 true properties (μ and σ) of one simulated batch. The 85% iso-coverage curve (the
 1373 limiting quality of the PTI tests) is shown to guide the eye. See text for details.

1374



1375

1376

1377

1378
1379
1380
1381
1382
1383

1384
1385

Figure 23. Comparison of two of the PTI tests described in the *Table of Test Plans* (including the $100 \pm 15\%$ LC requirement on the sample average for each life-stage separately), and the FDA DCU&TCL test in a simulated production situation for a multi-dose product. Each dot represents the true properties (μ and σ) of one simulated batch. The 85% iso-coverage curve (the limiting quality of the PTI tests) is shown to guide the eye. See text for details.

1385

1386

CONCLUSIONS

1387

1388

- 1389 • A parametric tolerance interval test (PTI test) is proposed as a replacement for the
1390 delivered dose uniformity tests in the FDA draft Guidances. The PTI test is more
1391 efficient because it makes more complete and thorough use of the information obtained
1392 from a sample. The ability to reliably estimate the quality of the batch from which the
1393 sample originates is improved, which increases the likelihood for a correct disposition
1394 of the batch.
- 1395 • For single-dose products, it has been demonstrated that compared to the FDA DCU test,
1396 the PTI test provides superior consumer protection against false acceptance of batches
1397 that do not fulfill the specified limiting quality requirement. Stated in statistical terms,
1398 the PTI test provides a higher coverage of the $100\pm 25\%$ LC target interval compared to
1399 the FDA DCU test (*i.e.*, minimum 85% coverage for the PTI test vs 78% coverage for
1400 the FDA DCU test, at the 95% confidence level).
- 1401 • For multi-dose products, it has been demonstrated that the PTI test provides consumer
1402 protection comparable to that given by the simultaneous application of the FDA DCU
1403 and TCL tests (*i.e.*, minimum 85% coverage of the $100\pm 25\%$ LC target interval, at the
1404 95% confidence level)
- 1405 • Single-dose as well as multi-dose products are proposed to be tested using a single PTI
1406 test. For multi-dose products, a stratified sampling plan is used, with equal testing of
1407 beginning, middle and end doses (or if appropriate, beginning and end doses) of
1408 different inhalers. The proposed stratified sampling plan allows simultaneous control of
1409 both between-container and through-container-life uniformity for multi-dose products.
1410 Similarly to the FDA TCL test, the mean of each tested life stage is required to be
1411 within $\pm 15\%$ of the label claim.
- 1412 • The proposed test replaces the zero tolerance limit of the FDA draft Guidance with a
1413 parametric limiting quality statement. As a result, the proposed PTI test provides
1414 protection against deviating doses comparable or superior to that given by the FDA
1415 draft Guidance tests.
- 1416 • The PTI test is applicable for normally as well as non-normally distributed data.
- 1417 • The PTI test provides superior protection, compared to the FDA draft Guidance tests,
1418 against false random rejections of batches of acceptable quality.
- 1419 • The proposed PTI test provides flexibility to the producer to choose a test plan that is
1420 most suitable for a particular product without compromising the consumer protection.
- 1421 • The cost incurred to achieve these advantages is an increased average demand on the
1422 sample size. As a result, the quality of each batch is more thoroughly investigated.
- 1423 • The proposed PTI test provides a powerful incentive for producers to improve their
1424 product quality since sample size requirements are lower for products of higher quality.

1425

1426

ACKNOWLEDGEMENTS

1427

1428

1429

1430

1431

1432

1433

1434

1435

1436

1437

1438

In January 2000, the International Pharmaceutical Aerosol Consortium on Regulation and Science (IPAC-RS) and scientists of the Inhalation Technology Focus Group (ITFG) of the American Association of Pharmaceutical Scientists (AAPS) initiated a scientific collaboration to address important issues in the FDA's draft Guidance documents for orally inhaled and nasal drug products, including the FDA's dose content uniformity specification. A DCU Working Group of leading pharmaceutical scientists was formed to investigate further this issue. The development of *A Parametric Tolerance Interval Test for Improved Control of Delivered Dose Uniformity of Orally Inhaled and Nasal Drug Products* is the product of the DCU Working Group's effort.

1439

1440

1441

1442

1443

1444

1445

1446

1447

The general approach of a parametric tolerance interval test was inspired by Dr. Walter Hauck of Thomas Jefferson University. Dr. Hauck provided valuable input to the Working Group as it developed the proposal. The statistical design of the test suggested by the Japanese Pharmacopeia and the Pharmacopeial Discussion Group of ICH (described in the recently published revised draft of USP Chapter <905> *Uniformity of Dosage Units*) served as another important source of inspiration. FDA's commitment to a scientific, data-driven discussion with industry has served as a driving force for the development of a scientifically sound test that will help ensure the availability of high quality orally inhaled and nasal drug products for patients.

1448

1449

1450

1451

1452

1453

1454

1455

1456

1457

From its inception, the DCU Working Group benefited from the outstanding leadership of three individuals: **Dr. Bo Olsson, Dr. Cynthia Flynn and Dr. Dennis Sandell**. Dr. Olsson, the principal architect of the DCU Working Group's proposal, for many years served as Scientific Advisor for Pharmaceutical and Analytical R&D at AstraZeneca; currently he is Director of Scientific and Regulatory Affairs at Microdrug Development AB in Lund, Sweden. Dr. Flynn until recently was the Director of Pharmaceutical Sciences at Aventis and is currently with R. W. Johnson Pharmaceutical Research Institute in Spring House, Pennsylvania. Dr. Sandell is a Principal Research Scientist at AstraZeneca in Lund, Sweden, and served as the lead statistician of the DCU Working Group.

1458

1459

1460

1461

1462

1463

1464

1465

1466

1467

1468

1469

We acknowledge the statisticians of the DCU Working Group: Dr. Kristi Griffiths of Eli Lilly, Dr. Sharon Murray of GlaxoSmithKline, Mr. Edward Warner of Schering-Plough, and Dr. Buffy Hudson-Curtis of GlaxoSmithKline and the other core members of the DCU Working Group: Mr. Mark Broughton of Aventis Pharmaceuticals, Mr. Michael Golden of GlaxoSmithKline, Dr. Igor Gonda of Aradigm, Dr. Paul Kovach of Eli Lilly, Dr. Stefan Leiner of Boehringer Ingelheim, Dr. John Morgan of GlaxoSmithKline, Dr. David Radspinner of Aventis Pharmaceuticals, Mr. Keith Truman of GlaxoSmithKline, Mr. Steve White of Inhale Therapeutic Systems, Dr. Bruce Wyka of Schering-Plough, and Dr. Svetlana Lyapustina, IPAC-RS Science Advisor. We also acknowledge all of the individuals who supported the work of the DCU Working Group and all of the companies that submitted DCU data.

1470

1471

1472

This effort has been sponsored by the IPAC-RS companies: Aradigm, Armstrong Pharmaceuticals, AstraZeneca, Aventis, Boehringer Ingelheim, Eli Lilly, GlaxoSmithKline, Inhale Therapeutic Systems, Kos Pharmaceuticals, IVAX, Pfizer, and Schering-Plough.

1473

1474

APPENDICES

1475

1476

1477

1 FDA Draft Guidance Test

1478

1.1 Interpretation of the FDA Tests

1479

The tests listed in the *Metered Dose Inhaler and Dry Powder Inhaler Drug Products CMC*

1480

Draft Guidance were interpreted as follows:

1481

1482

Dose Content Uniformity (Sections III.F.1.i and III.F.2.h)

1483

(This test is referred to as the FDA DCU test)

1484

1485

For each of ten containers, determine one dose. The test is passed if

1486

- NMT 1 of the 10 values is outside $\pm 20\%$ label claim (LC),

1487

- None is outside $\pm 25\%$ LC, and

1488

- The average of the 10 values is within $\pm 15\%$ LC.

1489

If the test is not passed, twenty additional containers are tested in a 2nd tier provided the 1st tier

1490

average is within 15% LC, NMT 3 values are outside $\pm 20\%$ LC, and no value is outside $\pm 25\%$

1491

LC. The test is passed if

1492

- NMT 3 of the 30 values are outside $\pm 20\%$ LC,

1493

- None is outside $\pm 25\%$ LC, and

1494

- The average of the 30 values is within $\pm 15\%$ LC.

1495

1496

Dose Content Uniformity Through Container Life (Sections III.F.1.j and III.F.2.i)

1497

(This test is referred to as the FDA TCL test)

1498

1499

For each of three containers, determine one beginning, one middle, and one end dose. The test

1500

is passed if

1501

- NMT 1 of the 9 values is outside $\pm 20\%$ LC,

1502

- None is outside $\pm 25\%$ LC, and

1503

- The average of each of the beginning, middle and end values are all within $\pm 15\%$ LC.

1504

If the test is not passed, six additional containers are tested in a 2nd tier provided all three 1st tier

1505

averages are within 15% LC, NMT 3 values are outside $\pm 20\%$ LC, and no value is outside

1506

$\pm 25\%$ LC. The test is passed if

1507

- NMT 3 of the 27 values are outside $\pm 20\%$ LC,

1508

- None is outside $\pm 25\%$ LC, and

1509

- The average of each of the beginning, middle and end values are all within $\pm 15\%$ LC.

1510

1511

For Nasal Spray and Inhalation Solutions, Suspension, and Spray Drug Products, the CMC

1512

draft Guidance recommends a *Spray Content Uniformity* test that is interpreted to be equivalent

1513

to the DCU test for MDI/DPI drug products. The *Spray Content Uniformity Through*

1514

Container Life test differs from the *Dose Content Uniformity Through Container Life* test in

1515

that one beginning spray and one end spray is sampled from each of 5 containers, otherwise the

1516

corresponding requirements apply.

1517

1518 1.2 Operating Characteristic Curves for the FDA DCU Test

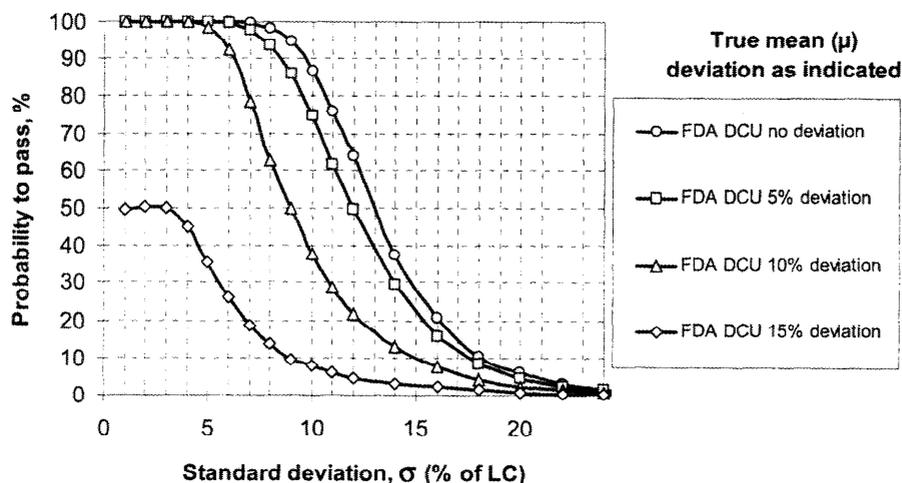
1519 The FDA DCU test is applied to single-dose products such as capsules and blisters. It is also
 1520 applied to multi-dose products, but then in combination with the FDA TCL test. This section
 1521 reviews the operating characteristics of the FDA DCU test used alone. Section 1.3 reviews the
 1522 simultaneous application of both the DCU and the TCL tests.

1523

1524 The operating characteristic (OC) curve for the FDA DCU test is given in Figure 24 for a
 1525 normal distribution with the true mean at the target (100 % LC). OC curves for a true mean
 1526 deviating by 5%, 10% and 15% LC from the target are also shown in Figure 24. For
 1527 distributions that deviate from the target, the OC curves are shifted towards smaller standard
 1528 deviations.

1529

1530



1531

1532 **Figure 24. Operating Characteristic curves for the FDA DCU test based on normal distribution**
 1533 **with batch mean deviation from target and batch standard deviation as indicated.**

1534

1535 Another way of illustrating the operating characteristics for the FDA Draft Guidance test is to
 1536 employ a graph of batch standard deviation *versus* batch mean. On such a graph, the batch
 1537 quality (as expressed by mean and standard deviation) corresponding to a 5% acceptance
 1538 probability (5% consumer risk) and 95% acceptance probability (5% producer risk) can be
 1539 outlined. Figure 25 (page 50) shows these iso-probability curves for the FDA DCU test. The
 1540 complex curve-form is derived from simultaneously applying the three acceptance rules in the
 1541 FDA DCU test:

1542

1543

1544

1545

1546

- Sample mean within 100±15% LC;
- 90% of the sample observations inside 100±20% LC (inner limits);
- No observation in a sample is outside 100±25% LC (outer limits).

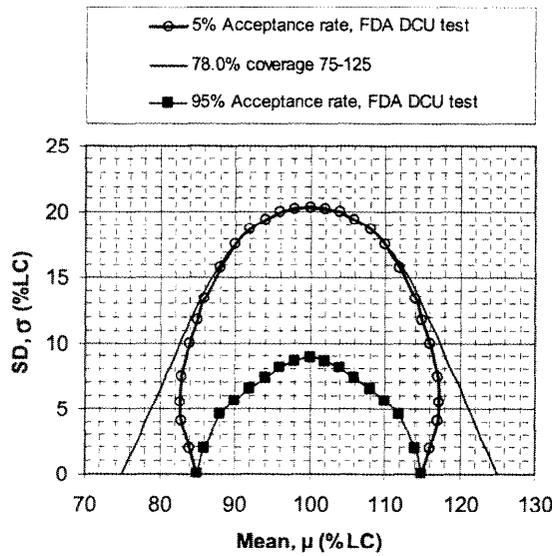
1547 Figure 26 to Figure 28 below show the iso-probability curves for each of these separate
1548 requirements. The upper curve labeled “5% acceptance rate” represents the quality of batches
1549 that have exactly 5% probability to provide a sample that passes the specified requirement of
1550 the test. The lower curve labeled “95% acceptance rate” represents the quality that has exactly
1551 95% probability to provide a sample that passes the specified requirement of the test.

1552
1553 The mean requirement of the FDA DCU test is seen to be responsible for control in the region
1554 of large deviations from the target (compare Figure 25 and Figure 26). For the inner limit
1555 requirement (Figure 27), the 5% iso-probability curve traces an iso-coverage curve of 60.5%
1556 coverage of the interval 80-120% LC. This means that with 95% probability, this component
1557 of the test will reject a batch that has 60.5% of the doses within 80-120% of the label claim.
1558 For the outer limit requirement (Figure 28), the 5% iso-probability curve traces an iso-coverage
1559 curve of 74% coverage of the interval 75-125% LC. This means that with 95% probability, this
1560 component of the test will reject a batch that has 74% of the doses within 75-125% of the label
1561 claim.

1562
1563 Because all three requirements are applied simultaneously, the iso-probability curve for the
1564 complete FDA DCU test (Figure 25) does not trace any particular iso-coverage curve.
1565 However, for batch means not too distant from the target (approximately $100 \pm 12\%$ LC), the
1566 iso-probability curve for 5% acceptance probability closely traces the iso-coverage curve for
1567 78%²⁴ coverage of the interval 75-125%²⁵ LC (compare the 78% iso-coverage curve and the
1568 5% iso-probability curve in Figure 25). This means that with 95% probability, the FDA DCU
1569 test will reject a batch that has 78% of the doses within the interval 75 –125% LC. For batch
1570 means further away from the target, the test becomes more conservative due to the requirement
1571 on the mean.
1572

²⁴ The reason that this coverage (78%) is slightly higher than that given by the outer requirement alone (74%, see Figure 28) is the complex interaction of the requirements with regard to 2nd tier testing.

²⁵ The 5% iso-probability curve also traces an iso-coverage curve representing a lower coverage (67%) of a tighter interval (80-120% LC) for batch means not too distant from the target (approximately $100 \pm 12\%$ LC). Hence, in this region of batch means, the two statements “78% coverage of 75-125% LC” and “67% coverage of 80-120%” are equivalent.



1573
1574

Figure 25. Iso-probability curves for the FDA DCU test (all requirements)

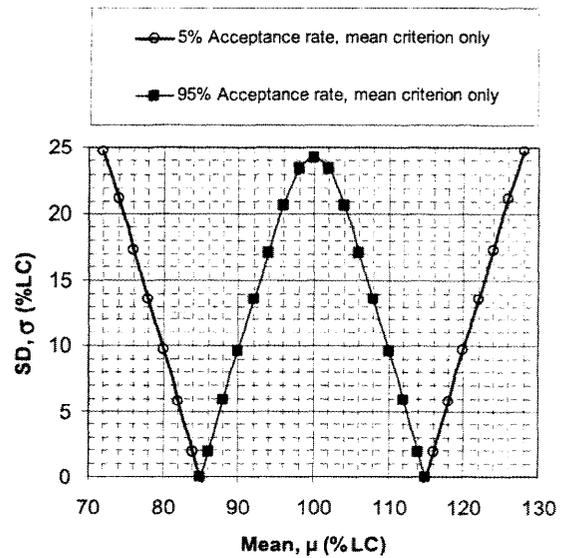
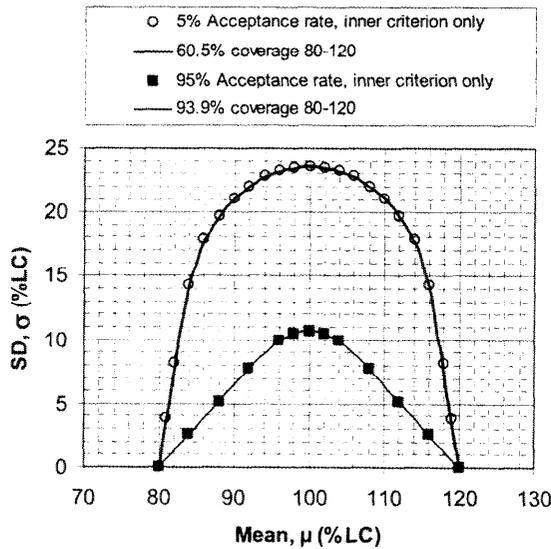


Figure 26. Iso-probability curves for the requirement on the mean, only

1575
1576



1577

Figure 27. Iso-probability and iso-coverage curves for the inner limits, only

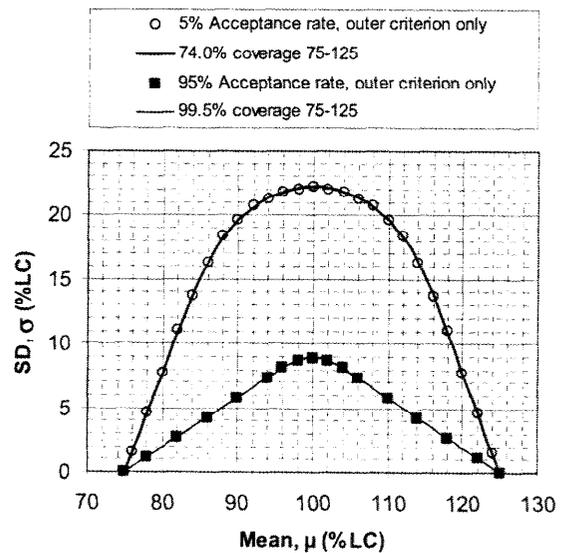


Figure 28. Iso-probability and iso-coverage curves for the outer limits, only

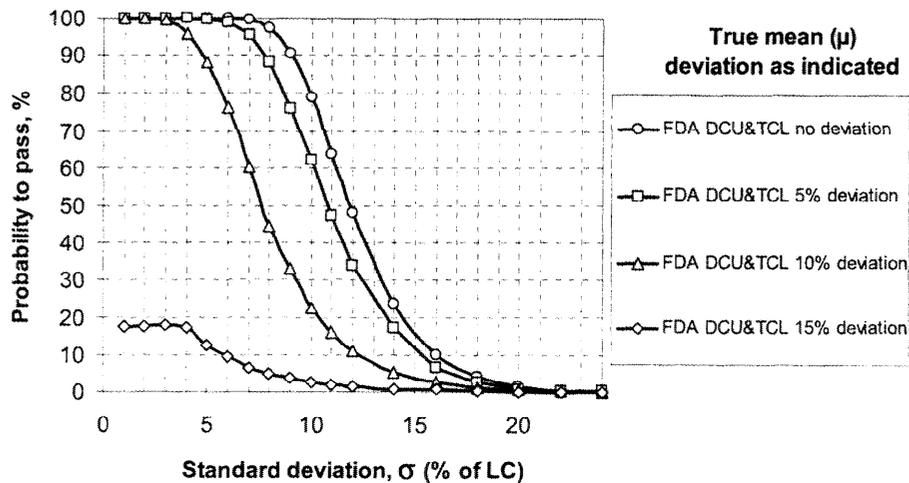
1578
1579

1.3 Operating Characteristic Curves for the Simultaneous Application of the FDA DCU and TCL Tests

For multi-dose products, such as pressurized metered dose inhalers, reservoir dry powder inhalers and multi-cavity blister pack inhalers, the draft Guidances prescribe that uniformity is tested by application of both the DCU and the TCL tests. These tests can be combined in a number of different ways. For the present purpose of investigating the operating characteristics of a combination of these two tests, the following assumptions were made:

- Ten beginning doses are sampled from ten inhalers for the DCU test;
- The first three inhalers used for the DCU test are also sampled in the middle and end of container life for the TCL test;
- The intra- and inter-inhaler variability are of equal magnitude.

The OC curve for the FDA DCU&TCL test is given in Figure 29 for a normal distribution with the true mean at the target (100 % LC). OC curves for a true mean deviating by 5%, 10% and 15% LC from the target are also shown in Figure 29.



1598

Figure 29. Operating Characteristic curves for the simultaneous application of the FDA DCU and TCL test, based on normal distribution with batch mean deviation from target and batch standard deviation as indicated.

Comparing the OC curves for the FDA DCU&TCL tests (Figure 29) with the OC curves for the FDA DCU test (Figure 24), it is apparent that the acceptance rate is consistently lower for the combined tests (given any quality). This is due to the fact that in order to be accepted, both tests must be passed. For low quality multi-dose products (high standard deviation and/or large mean deviation), this provides added consumer protection compared to single-dose products. For high quality multi-dose products (low standard deviation and small mean deviation), however, this increases significantly the probability to fail a batch due to a random observation, *i.e.*, the producer risk is increased.

1611 As demonstrated in Figure 30, the simultaneous application of the DCU and TCL tests results
 1612 in an acceptance rate of 5% for a coverage of 85% of the $100\pm 25\%$ LC interval (for mean
 1613 deviations within $\pm 12\%$ of the target). Compared to the coverage afforded by the FDA DCU
 1614 test alone (78%), this demonstrates that the quality requirements in the draft Guidances are
 1615 more stringent for multi-dose than for single-dose products.

1616

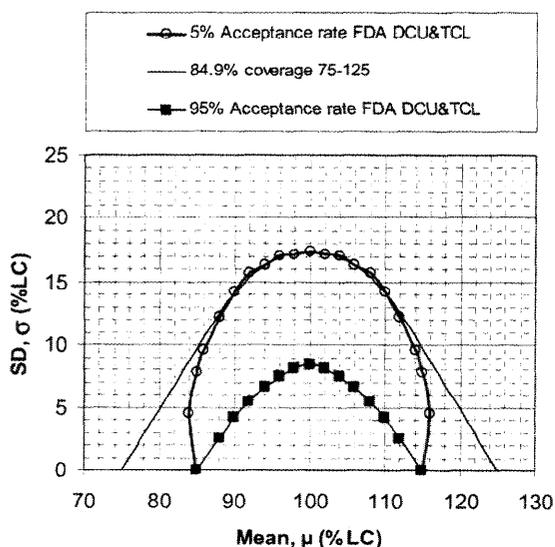
1617 For multi-dose *sprays*, simultaneous application of the *Spray Content Uniformity* and the *Spray*
 1618 *Content Uniformity Through Container Life* tests results in a limiting coverage of 84% (using
 1619 similar assumptions as those stated above for DCU&TCL testing).

1620

1621 The limiting coverage implied by the DCU&TCL tests is slightly affected by the assumptions
 1622 made in calculating the operating characteristics. Depending on particular assumptions
 1623 regarding the relation between intra- and inter-inhaler variability and how the observations are
 1624 combined for the two tests, the limiting coverage implied by the FDA tests varies between 78%
 1625 and 88%. The assumptions listed in the beginning of this section, which result in a coverage of
 1626 85%, are judged to be appropriate and realistic.

1627

1628



1629

1630 **Figure 30. Iso-probability curves for the simultaneous application of the FDA DCU and TCL**
 1631 **tests.**

1632

1633 The above analyses of the FDA DCU test and the FDA DCU&TCL tests were performed in
 1634 order to establish a baseline for the development of a new, more efficient test that would
 1635 provide the same or a better level of limiting quality as the FDA draft Guidance tests.

1636

1637 Given that the FDA DCU test (for a single-dose product) provides a limiting (*i.e.*, at 5%
 1638 acceptance rate) coverage of 78% of the 75-125% LC interval, and that the FDA DCU&TCL
 1639 tests (for a multi-dose product) provides a limiting coverage of 85%, it was assumed that a
 1640 replacement test, aiming to be applicable for both cases, needs to provide a minimum coverage
 1641 of 85% (there is no reason to allow a lower limiting coverage for single-dose products than for

1642 multi-dose products). The proposed PTI test was designed to have a 5% consumer risk of
1643 accepting a batch that has an 85% coverage of a 75-125% LC target interval. Thus, the PTI test
1644 provides higher batch coverage for the same 5% consumer risk for a single-dose product and
1645 equally high coverage for a multi-dose product.

1646

1647 Alternatively stated, with the proposed PTI test, a batch with 78% coverage of the 75-125% LC
1648 target interval has less than 5% acceptance probability, and the consumer risk is thus lower for
1649 a single-dose product with the PTI test than with the FDA DCU test (which would accept such
1650 a batch with 5% probability). For a multi-dose product, the high consumer protection provided
1651 by the FDA DCU%TCL tests is matched by the PTI test.

1652

1653 **1.4 Factors Controlling the Outcome of the FDA Tests**

1654 The operating characteristics of the draft Guidance tests were investigated by simulations,
1655 using the following models.

1656

1657 For a single-dose product, it was assumed that observations were normally distributed with the
1658 mean at target. In each round, ten values were randomly selected from the distribution,
1659 representing one dose from each of ten containers, and the FDA DCU test was applied. Second
1660 tier testing and evaluation followed the rules given in the draft Guidance as interpreted above.
1661 The proportion of non-complying samples and the cause for non-compliance are given in Table
1662 2 (the rows labeled "DCU").

1663

1664 For a multi-dose product, it was assumed that the observations were normally distributed with
1665 the mean at target, that no through-container-life trend existed, and that the overall variation
1666 emanated from both inter- and intra-inhaler variability (of equal magnitude). In each round, ten
1667 inhalers were randomly selected from the distribution of inhalers. From each of these, one
1668 dose was randomly drawn from the distribution of doses, and the FDA DCU test was applied.
1669 From three of these inhalers, a further two doses were drawn to represent middle and end
1670 doses. The nine doses from these three inhalers were subjected to the TCL test. For each test,
1671 second-tier testing was performed according to the rules for the respective test. Evaluation
1672 followed the rules given in the draft Guidance as interpreted above. The proportion of non-
1673 complying samples, and the cause for non-compliance are given in Table 2 (all rows).

1674

1675 Note that the models above for single- and multi-dose products yield exactly the same result for
1676 the DCU part of the testing (which is the complete test for a single-dose product).

1677

1678 For a batch with the true mean at target, the overall rate of non-compliance with the FDA DCU
1679 test (single-dose product) was found to be 1.8, 13.1, and 37.3% for a true batch standard
1680 deviation of 8, 10, and 12 % LC, respectively (Table 2). For the FDA DCU&TCL test (multi-
1681 dose product) the corresponding rates of non-compliance were 3.0, 20.3, and 51.7%. In all
1682 cases, the absolute majority of failures was caused by obtaining a value outside the outer limits.
1683 Yet, for these distributions, the true proportion of values outside $100 \pm 25\%$ LC is 0.2, 1.2 and
1684 3.7%, respectively (*i.e.*, the true coverage is 99.8, 98.8 and 96.3%, respectively). Also note that
1685 the mean requirements of the TCL test lead to a number of rejections, even though the true
1686 mean is on target and no through-container-life trend is present.

1687

1688

1689
1690
1691
1692
1693

Table 2. Rate (%) and cause of non-compliance with FDA draft Guidance tests assuming normal distribution, true batch mean at 100% LC and true overall batch standard deviation (σ , % LC) as indicated. For a single-dose product, only the part labeled DCU applies. For a multi-dose product, the complete table applies.

<i>Test</i>	<i>Cause for non-compliance</i>	$\sigma=8.0$	$\sigma=10.0$	$\sigma=12.0$
DCU	Failed inner limits*	0.0	0.9	6.1
	Failed outer limits*	1.8	12.7	35.7
	Failed inner and/or outer limit*	1.8	13.1	37.3
	Failed mean*	0.0	0.0	0.0
	Failed any DCU criteria*	1.8	13.1	37.3
	Total allowed into 2 nd tier	0.5	3.6	8.1
TCL	Failed inner limits*	0.0	1.1	5.1
	Failed outer limits*	1.5	10.1	28.3
	Failed inner and/or outer limit*	1.6	10.6	29.5
	Failed mean*	0.4	2.6	7.6
	Failed any TCL criteria*	1.8	11.8	31.8
	Total allowed into 2 nd tier	0.8	3.1	5.8
DCU & TCL	Total failed 2 nd tier (any cause)	0.1	1.0	3.7
	Failed any DCU and/or TCL criteria*	3.0	20.3	51.7

1694
1695
1696

* 2nd tier failures are included

1697
1698
1699
1700

Second tier testing is not permitted if the first tier testing results in a value outside the outer limit requirements. This fact results in the low incidence of the 2nd tier testing as demonstrated in Table 2.

1701
1702
1703
1704
1705
1706
1707

The results of these simulations demonstrate that it is the zero-tolerance requirement for the outer limits that to all practical purposes determines the outcome of the FDA tests. It is further demonstrated that for the same quality (coverage), a multi-dose product fails more frequently than a single-dose product. Stated in a different way, this demonstrates that the FDA draft Guidance tests require a higher uniformity for multi-dose products than for single-dose products.

1708
1709
1710
1711
1712
1713
1714
1715
1716

The critical role of the outer limits requirement is further highlighted by the imbalance between the requirement that no value is allowed outside 100±25% LC and the allowance of up to 10% of values outside ±20%, as shown for a normal distribution in Table 3. The table demonstrates that for a distribution with a standard deviation of 12.2% LC, where 1 out of 10 (*i.e.* 10%) values are outside the 100±20% LC interval, 4% of the values are outside the 100±25% LC interval. This explains why the requirement that no observation be allowed outside the outer limits determines the outcome of the test.

1717 **Table 3. Relation between inner and outer limits (assuming that data are normally**
 1718 **distributed with a mean of 100% LC)**

Standard Deviation, % LC	% Outside 100±20% LC	% Outside 100±25% LC
7.1	0.5	0.05
7.8	1	0.1
10.2	5	1.4
12.2	10	4.0

1719

1720

1721 Because some fraction (however small) of the doses in a batch is always outside the outer
 1722 limits, the probability of failing the requirement that *none* may be observed increases steadily
 1723 with the number of observations. Thus, the more thorough the investigation, the more certain
 1724 is a failure. This is in contrast to the requirement that *not more than 10%* may be outside the
 1725 inner limits; here the more thorough the investigation, the more certain is a correct decision. In
 1726 Table 4, some examples showing the probability to observe a value outside the outer limits are
 1727 given for different true fractions outside the outer limits and different number of observations.

1728

1729

1730

Table 4. Risk to fail (%) outer limits for different number of observations^{*}.

True fraction (%) outside outer limits, 100q	Corresponding Coverage (%)	Number of observations, n					
		10	16	100	160	300	480
0.1	99.9	1	2	10	15	26	38
0.3	99.7	3	5	26	38	59	76
1.0	99.0	10	15	63	80	95	99
3.0	97.0	26	39	95	99	100	100

1731

* calculated from $1-(1-q)^n$

1732

1733

1734 Typically, a stability investigation of three batches would involve about 300 observations (or
 1735 480 observations for a multi-dose product). Table 4 shows, for example, that for a hypothetical
 1736 product with 0.1% values outside the outer limits (which for a normal distribution with the
 1737 mean at target corresponds to 1% values outside the inner limits; compare Table 3) there is a
 1738 26% probability to obtain at least one such value if 300 values are observed (38% probability
 1739 for 480 observations), while if only 10 (16) values are observed the risk is reduced to 1% (2%),
 1740 although the quality of the product *is the same*.

1741

1742 This constitutes a strong incentive for the industry to minimize the number of observations, *i.e.*
 1743 the FDA zero tolerance criterion rewards small investigations and penalizes thorough ones. In
 1744 any kind of investigation where many observations are called for (such as development,
 1745 stability, validation), this issue will need to be addressed.

1746

1747

1748

2 Distribution of Data in OINDP Database

1749

1750

1751

1752

1753

1754

1755

1756

1757

1758

1759

1760

1761

1762

1763

1764

1765

1766

1767

1768

1769

1770

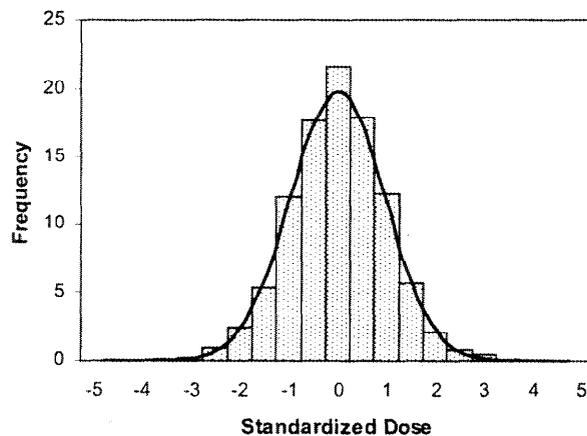
1771

1772

The proposed PTI test was developed assuming that delivered dose uniformity data follow a normal distribution. It is therefore important to investigate whether the assumption of a normal distribution is valid.

This appendix uses the ITFG/IPAC-RS DCU database to illustrate the actual distribution of delivered dose data, and compare this to the normal distribution.

The ITFG/IPAC-RS DCU database contains data for 80 products and a total of 46,816 individual determinations. In Figure 31, a histogram showing the overall distribution of these data is presented. In the same figure, the density function for the standardized normal distribution is superimposed for comparison. Note that in order to be able to pool all data from different products and present an overall summary, the data needed to be standardized. Consider for example a situation where data is available for two batches, both following a normal distribution but with different means (say, 90 and 110% LC). If these data are combined and displayed in one histogram, a distribution indicating non-normality (a bi-modal one) would be obtained. A similar artifact could be obtained due to differences in variability. To address these concerns, the following approach was used to construct Figure 31. For each product, the overall mean (m) and standard deviation (s) were calculated, after which each individual result was standardized $[(\text{dose}-m)/s]$ to the mean and standard deviation for the corresponding product. With this approach, all products are standardized to a mean of zero and a standard deviation of unity, allowing the data to be pooled in order to illustrate the general *shape* of the distribution.



1773

1774

1775

Figure 31. Histogram over standardized delivered dose data for all products (n=46,816), compared to the standardized normal distribution.

1776

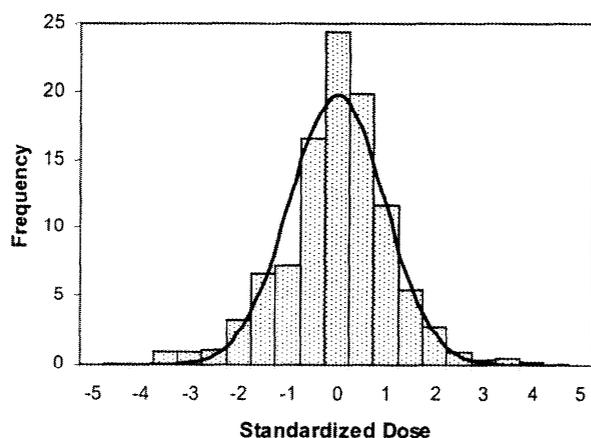
1777 Figure 31 shows that the overall distribution of data is symmetrical and that the fit to the
1778 standardized normal distribution is superb. Note specifically that the excellent fit extends far
1779 out into the tails of the distribution.

1780

1781 In addition to the overall illustration for all products in the database, the normality of the
1782 distribution of several individual products was investigated. These were selected to include a
1783 product from each of the main product types to assess potential differences between product
1784 types. Standardized histograms for one CFC MDI, one HFA MDI, one pre-metered DPI, one
1785 device-metered DPI, and one non-pressurized nasal spray are presented and compared to the
1786 standardized normal distribution in Figure 32 through Figure 36 below. For each type, the
1787 product with the largest number of available observations was selected.

1788

1789



1790

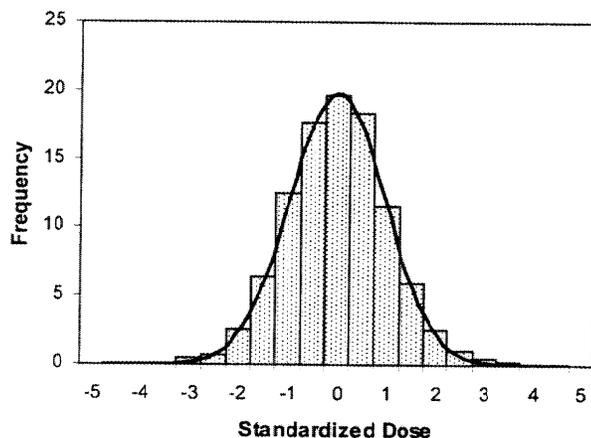
1791 **Figure 32. Histogram over standardized delivered dose data for a CFC MDI (n=1,310), compared**
1792 **to the standardized normal distribution.**

1793

1794 Data for the selected CFC MDI product (1,310 determinations from 9 batches) is shown in
1795 Figure 32. The right-hand side appears to follow the theoretical normal distribution to a high
1796 degree of accuracy, while some irregularities can be seen in the left part.

1797

1798



1799

1800 **Figure 33. Histogram over standardized delivered dose data for an HFA MDI (n=2,230),**
1801 **compared to the standardized normal distribution.**

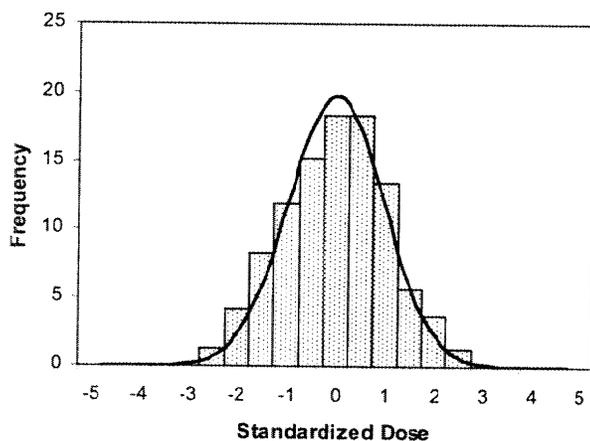
1802

1803 The HFA MDI product presented in Figure 33 (2,230 determinations from 6 batches) shows an
1804 excellent fit to the normal distribution.

1805

1806

1807



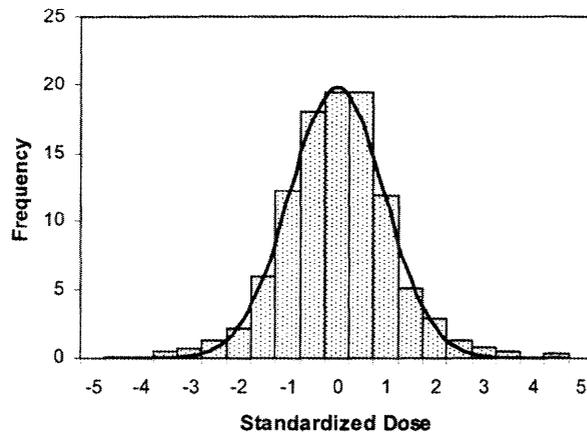
1808

1809 **Figure 34. Histogram over standardized delivered dose data for a pre-metered DPI (n=200),**
1810 **compared to the standardized normal distribution.**

1811 The limited available data (only 200 determinations from 3 batches) for the pre-metered DPI in
1812 Figure 34 follow the normal distribution very well.

1813

1814



1815

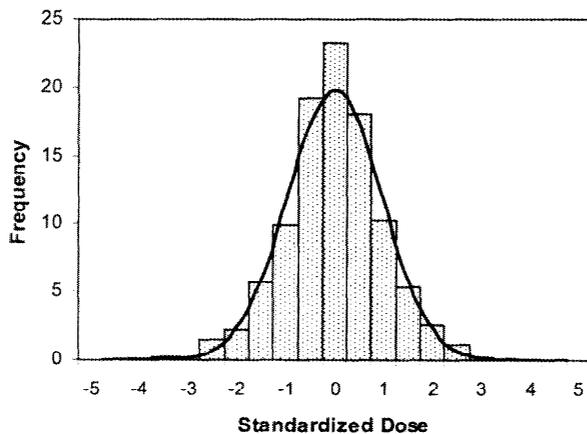
1816 **Figure 35. Histogram over standardized delivered dose data for a device-metered DPI (n=3,658),**
 1817 **compared to the standardized normal distribution.**

1818

1819 In Figure 35, data for a device-metered DPI is presented (3,658 determinations from 18
 1820 batches). This product displays an excellent fit to the normal distribution.

1821

1822



1823

1824 **Figure 36. Histogram over standardized delivered dose data for a non-pressurized nasal spray**
 1825 **(n=1,200), compared to the standardized normal distribution.**

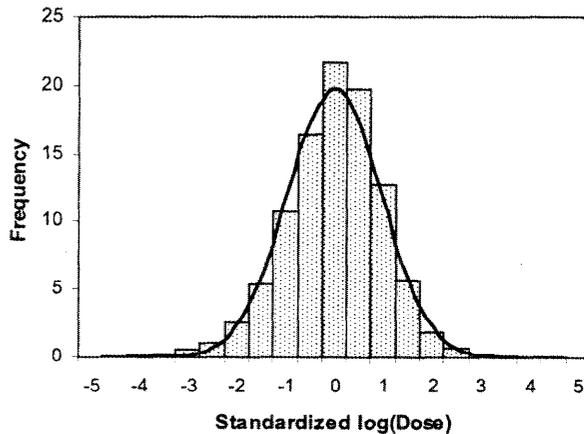
1826

1827 In Figure 36, data for a non-pressurized nasal spray is presented (1,200 determinations from 32
 1828 batches). This product also displays an excellent fit to the normal distribution.

1829

1830 For completeness, it was investigated whether the data could be fitted to a log-normal
 1831 distribution²⁶. To study this alternative, all individual determinations were log-transformed and
 1832 a plot corresponding to Figure 31 was constructed. This is shown in Figure 37 below.

1833
 1834



1835

1836 **Figure 37. Histogram over standardized log-transformed delivered dose data (n=46,816),**
 1837 **compared to the standardized normal distribution.**

1838

1839 Figure 37 shows that the log-normal assumption also provides a good fit to the data, although
 1840 the transformed distribution is slightly skewed to the right. By comparing the results of Figure
 1841 31 and Figure 37, it is concluded that the normal distribution provides a marginally better fit to
 1842 the data. In addition, as shown in Appendix 3, Section 3.2, the outcome of the PTI test is
 1843 virtually the same for a log-normal and a normal distribution having the same coverage.
 1844

1845 In summary, based on the delivered dose data collected by ITFG/IPAC-RS, it has been
 1846 demonstrated that the data are well approximated by a normal distribution, both overall and for
 1847 individual products of different types. This does not exclude the possibility that products exist
 1848 for which the normal assumption is not the best choice. However, results in Appendix 3 show
 1849 that the proposed test PTI is appropriate also in such instances, because the consumer
 1850 protection is not compromised.
 1851

²⁶ X is log-normal distributed if $Y=\log(X)$ is normal distributed.

1852

1853 **3 Applicability of PTI Test for Non-normally** 1854 **Distributed Data**

1855 The proposed Parametric Tolerance Interval (PTI) test has been developed based on the
1856 assumption that dose delivery data follow a normal distribution. Although the extensive
1857 collected database indicates that data for OINDP products typically are normally distributed
1858 (see Appendix 2), it cannot be ruled out that products exist for which this assumption is not
1859 fulfilled.

1860

1861 For this reason, it is important to investigate the performance of the PTI test for different
1862 potential deviations from normality, to ensure that the improved control provided by the
1863 proposed test is not degraded in situations where non-normal data are evaluated.

1864

1865 Three main types of deviations from normality have been studied to illustrate different
1866 situations of potential interest:

1867

1868 1. Multi-modal distributions.

1869

1869 2. Skewed distributions.

1870

1870 3. Heavy-tailed distributions.

1871

1872 The properties of the PTI test are studied below for each of these three potential situations. The
1873 general capability of the PTI test is demonstrated by presenting OC curves for the smallest
1874 ($n=10/30$) and largest ($n=24/72$) of the test plans given in the *Table of Test Plans* (page 7).
1875 Because the PTI test behaves similarly for single- and multi-dose products, the investigation
1876 focused on the tests for single-dose products without loss of generality.

1877

1878 **3.1 Multi-modal Distributions**

1879 A multi-modal distribution is a potential deviation from normality which can arise for different
1880 reasons (a bimodal distribution could occur, for example, when a fill weight adjustment is
1881 made during batch manufacture). It is important to ensure that the PTI test does not reward
1882 such undesirable characteristics; that is, for a fixed coverage the acceptance probability should
1883 not increase if data follow a multi-modal distribution.

1884

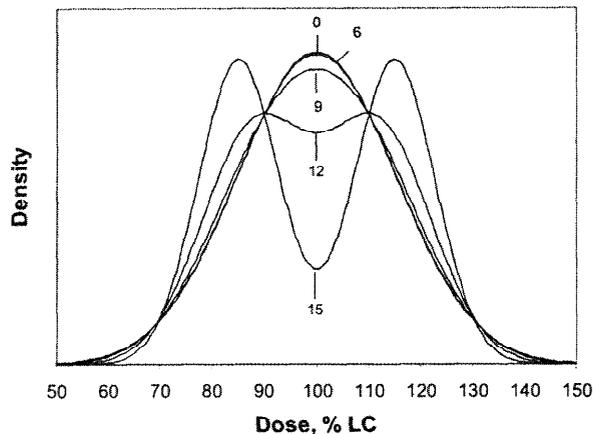
1885 As an example of a multi-modal distribution, bimodality was studied because this represents
1886 the worst case (with increasing number of modes, a multi-modal distribution tends to become
1887 less non-normal).

1888

1889 In Figure 38, the density of a normal distribution is compared to the densities of different
1890 bimodal distributions. All distributions in the figure have an overall mean at target (100% LC)
1891 and the same coverage of the target interval. The distance between the modes of the
1892 distribution ranges from $\pm 6\%$ (94 and 106% LC modes, representing a very slight non-
1893 normality that is barely distinguishable from a perfectly normal distribution) to $\pm 15\%$ (85 and
1894 115% LC modes), where separation of the modes is almost complete.

1895

1896 The bimodal distributions have been constructed by equally mixing two off-target normal
 1897 distributions having the same standard deviation. The coverage of the target interval (75-125%
 1898 LC) was varied between 80% and 99.999% by adjusting the standard deviation.
 1899
 1900



1901

1902 **Figure 38. Density for normal distribution ($d=0$) compared to bi-modal densities with modes at**
 1903 **$100 \pm d\%$ LC, $d = 6, 9, 12,$ and 15 .**

1904

1905 In Figure 39, OC curves for the PTI tests are presented for each of the five distributions. Note
 1906 that it is the coverage that is displayed on the horizontal axis rather than the standard deviation,
 1907 unlike in figures for OC curves in Part 2 of this report. This is necessary since when
 1908 comparing different *types* of distributions, there is no one-to-one correspondence between
 1909 coverage and standard deviation. For example, both the normal and the most extreme bi-modal
 1910 distribution in Figure 38 have the mean at target and the same coverage, but the overall
 1911 standard deviation is 15.2% and 16.9%, respectively. For this reason, an OC curve with
 1912 coverage rather than standard deviation on the horizontal axis has been used to compare the
 1913 distributions.

1914

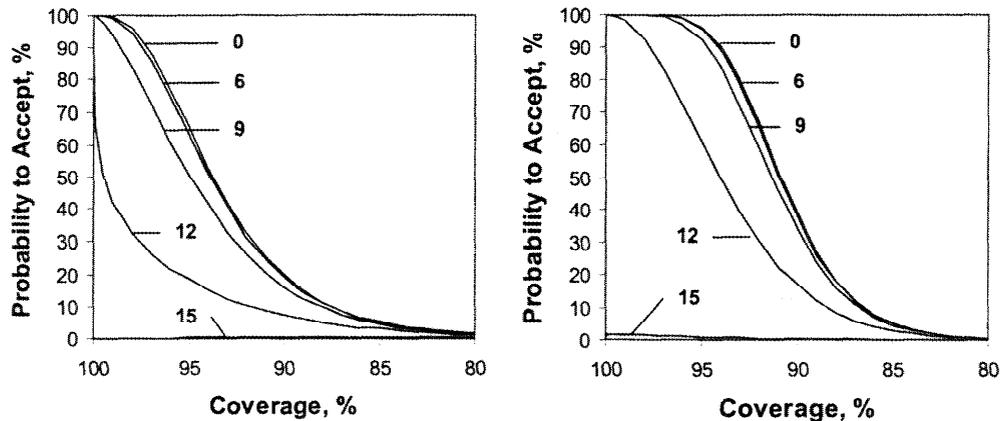
1915 Figure 39 shows that regardless of the degree of bimodality, the acceptance probability for any
 1916 of the PTI tests is always lower for a bimodal distribution than for the unimodal normal
 1917 distribution. Thus, there is no situation in which the presence of bimodality compromises
 1918 consumer protection.

1919

1920 Further, Figure 39 shows that the PTI tests control distributions with modes separated by less
 1921 than approximately $\pm 6\%$ to about the same degree as normally distributed data, while for larger
 1922 separations, the acceptance probability is significantly reduced. This provides a strong
 1923 incentive for manufacturers to avoid this kind of deficiency. Note that for products for which
 1924 data follows a distribution with $100 \pm 15\%$ modes, there is virtually zero probability of
 1925 acceptance.

1926

1927



1928

1929
1930
1931

Figure 39. OC curve for normal distribution ($d=0$) compared to those for bi-modal distributions with modes at $100 \pm d\%$ LC, $d = 6, 9, 12, 15$ (overall mean at target). PTI tests using $n=10/30$ (left panel) and $n=24/72$ (right panel).

1932

1933 3.2 Skewed Distributions

1934 A distribution of doses may potentially be asymmetrical. For example, if a multi-dose product
1935 with a through-container-life trend shows different variability for beginning and end doses, a
1936 skewed distribution may arise.

1937

1938 As an example of a skewed distribution, a shifted gamma distribution was chosen, with
1939 parameters selected to represent increasing degrees of deviation from normality. A random
1940 variable X is gamma distributed with parameters n and λ [denoted as $X \in \Gamma(n, \lambda)$] if it has the
1941 following density function:

1942

$$\frac{\lambda^n}{\Gamma(n)} x^{n-1} e^{-\lambda x}, x \geq 0.$$

1943

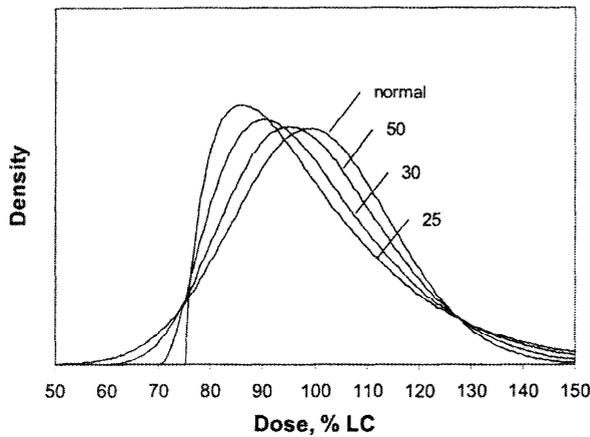
1944 Here, $\Gamma(n)$ denotes the gamma function (see the Glossary for definition). The mean of X is
1945 $\theta = n/\lambda$. For small values of θ , the distribution is skewed, while for larger θ values, it is fairly
1946 symmetric (in fact, X tends to a normal distribution when θ approaches infinity). To
1947 investigate a distribution that has the mean at target (100% LC) and which is skewed, a
1948 "shifted" gamma distribution $Y = (100 - \theta) + \Gamma(\theta, \lambda)$ is studied. The mean of Y is always at
1949 100, and increased skewness is obtained by decreasing θ . The desired coverage can be
1950 obtained by adjusting λ .

1951

1952 In Figure 40, the density functions for three different gamma distributions ($\theta = 25, 30, 50$) are
1953 compared to that of a normal distribution (all with the true mean at target and with the same
1954 coverage).

1955

1956



1957

1958
1959

Figure 40. Density for normal distribution compared to gamma densities with increasing degree of skewness ($\theta = 25, 30, 50$).

1960

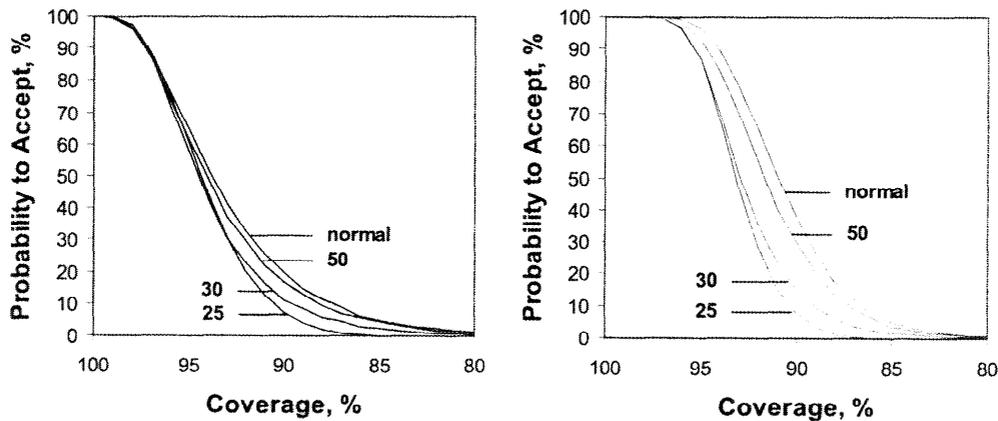
1961 The most extreme of the gamma distributions ($\theta = 25$) presented above is a very skewed
1962 distribution with a sharper mode than the normal distribution. The intermediate case ($\theta = 30$)
1963 also shows a marked difference compared to the normal case. When $\theta = 50$, only marginal
1964 differences from the normal distribution can be seen.

1965

1966 In Figure 41, OC curves for the PTI tests are presented for each of the four distributions. The
1967 coverage was varied between 80% and 99.5% by adjusting λ .

1968

1969



1970

1971
1972
1973

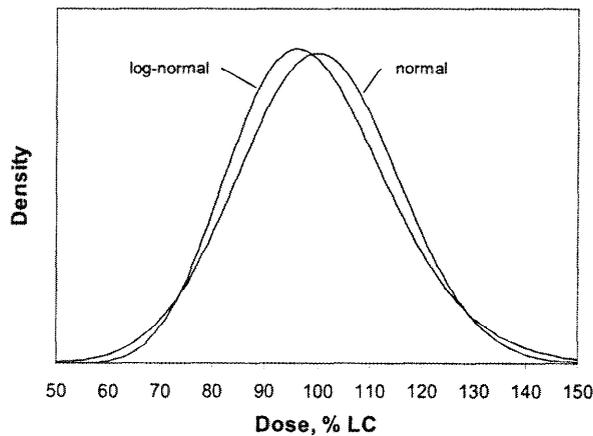
Figure 41. OC curve for normal distribution compared to those for gamma distributions with increasing degree of skewness, $\theta = 50, 30$ and 25 (mean at target). PTI tests using $n=10/30$ (left panel) and $n=24/72$ (right panel).

1974

1975 For the largest test in Figure 41 (right panel), the acceptance probabilities are consistently
 1976 smaller for the studied gamma distributions compared to the normal distribution. For the small
 1977 test (left panel), the acceptance rates for the gamma distributions are essentially the same as for
 1978 the normal distribution for high coverages (above about 95%), and are reduced compared to the
 1979 normal distribution for lower coverages.

1980
 1981 An interesting special case of a skewed distribution is the log-normal distribution (X is log-
 1982 normal distributed if $\log(X)$ is normal distributed). The special interest in this distribution
 1983 arises from the fact that delivered dose data sometimes is assumed to be log-normal.
 1984

1985 In Figure 42, the density functions for a log-normal distribution is compared to that of a normal
 1986 distribution (both with the true mean at target and with the same coverage).
 1987
 1988



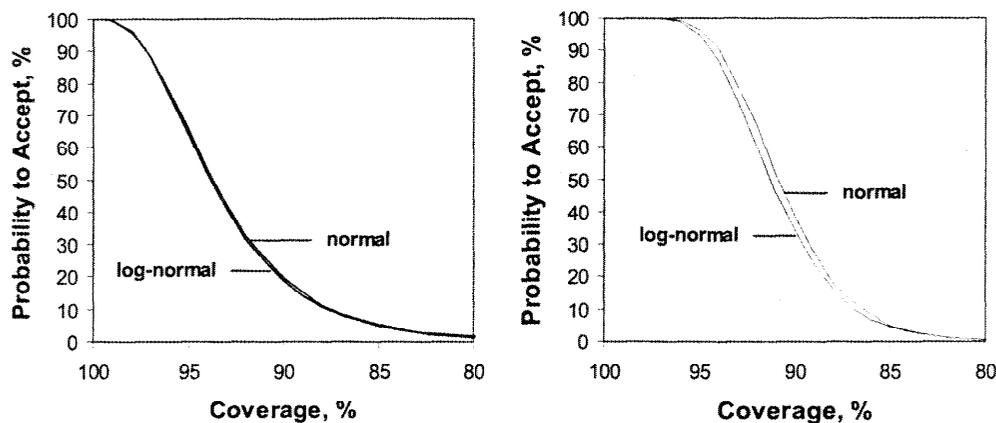
1989

1990 **Figure 42. Density for log-normal distribution compared to normal distribution.**

1991 The log-normal distribution represents a minor skewness to the left, with marginal differences
 1992 from the normal distribution.

1993

1994 In Figure 43, OC curves for the two PTI tests are presented for each of the two distributions.
 1995 The coverage was varied between 80% and 99.5% by adjusting the standard deviation.
 1996
 1997



1998

1999
2000

Figure 43. OC curve for normal distribution compared to log-normal distribution (mean at target). PTI tests using $n=10/30$ (left panel) and $n=24/72$ (right panel).

2001

2002

2003

2004

2005

2006

For the larger test in Figure 43 (right panel), the acceptance probabilities are consistently slightly lower for the log-normal distribution. For the smaller test (left panel), the acceptance probabilities are virtually identical for the two distribution types. These results indicate that the choice between modeling data by either of these distribution types has little practical relevance.

2007

3.3 Heavy-tailed Distributions

2008

2009

2010

2011

2012

2013

2014

2015

2016

For a heavy-tailed distribution, both small and large deviations from the mean are more common than for a normal distribution, whereas medium deviations are less common. A heavy-tailed distribution may arise, for example, if a drug formulation is inhomogeneous in such a way that the drug substance mixture contains rare agglomerates of the active drug in addition to the intended formulation. In such a case, the error would be non-negative and a small error would be likely more common than a large one. This may be modeled by disturbing a proportion of the values of a normal distribution by the addition of an exponentially distributed error.

2017

2018

2019

2020

2021

2022

The performance of the PTI tests in the presence of exponential disturbances is compared to that of the FDA tests in Part 2 of the report (Section 6.3). In this appendix, for completeness, non-normality caused by exponential disturbances is investigated using the same approach as used for the other types of non-normality (*i.e.*, comparing the acceptance rate for a normal and non-normal distributions having equal coverage).

2023

2024

2025

2026

In Figure 44, the density of an undisturbed normal distribution is compared to the densities of normal distributions disturbed to an increasing degree. All distributions in the figure have an overall mean at target and the same coverage.

2027

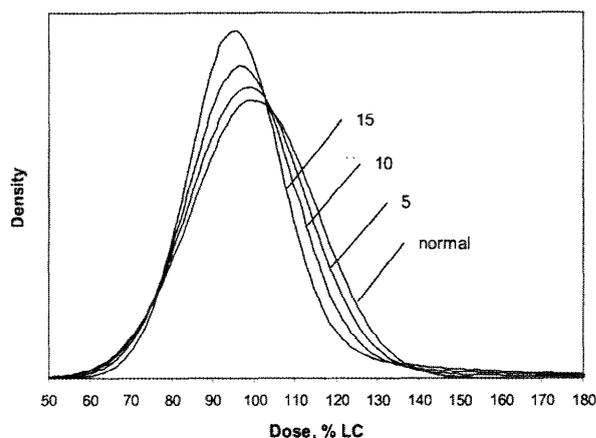
2028

2029

2030

The disturbed distributions have been constructed by adding an exponential error to the basic normal distribution: $[(1-p)N(100-\lambda p, \sigma) + p(N(100-\lambda p, \sigma) + \text{Exp}(\lambda))]$, where p is the (small) proportion of disturbed values and σ is adjusted to obtain the desired coverage. Here, $N(\mu, \sigma)$ denotes a normal distribution with mean μ and standard deviation σ , and $\text{Exp}(\lambda)$ an exponential

2031 distribution with parameter λ . (Note that the mean of the basic normal distribution is adjusted
 2032 to retain the overall mean of the disturbed distribution at target). The parameter λ , which for
 2033 an exponential distribution equals both the mean and the standard deviation, has been set to
 2034 35% LC. The resulting distributions are asymmetrical and have heavier right-hand tails and
 2035 sharper peaks (shifted to the left) than a pure normal distribution.
 2036
 2037



2038

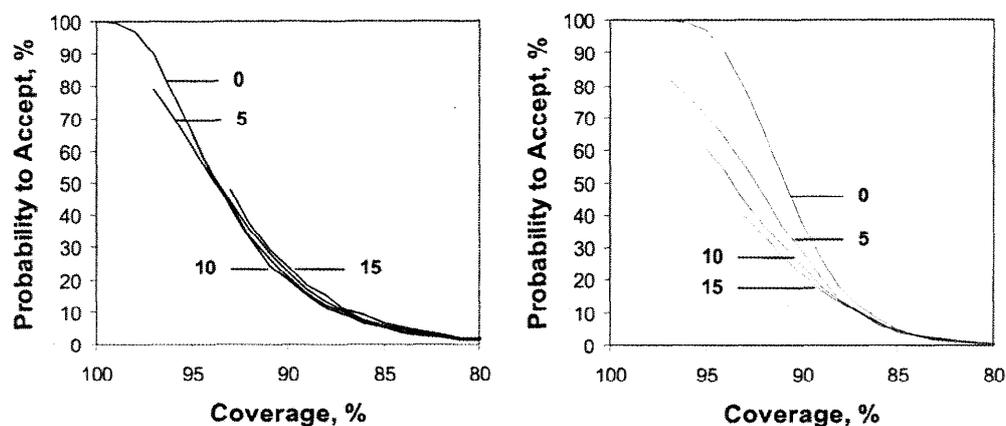
2039 **Figure 44. An undisturbed normal distribution compared to normal distributions affected by an**
 2040 **increasing proportion ($p=5, 10, 15\%$) of exponential disturbances (disturbance level,**
 2041 **$\lambda=35\%$ LC).**

2042

2043 In Figure 45, OC curves for the PTI tests are presented for the normal distribution and for the
 2044 three distributions affected by increasing proportions of exponentially disturbed values.
 2045

2046 The right-hand panel of Figure 45 shows that the acceptance probability for the larger PTI test
 2047 decreases significantly for high quality products (high coverages) when the normal distribution
 2048 is disturbed by an exponential distribution. For the smaller test (left panel), the effect is less
 2049 pronounced. For low quality products (low coverages), the acceptance probability for both PTI
 2050 tests is approximately the same as for an undisturbed normal distribution with the same
 2051 coverage. In particular, for the limiting coverage of 85%, the acceptance probability is
 2052 essentially constant at 5%. This indicates that when the PTI test is challenged by a heavy-
 2053 tailed distribution, the consumer protection is not degraded.
 2054

2055 The PTI test with the larger sample size clearly is more sensitive to disturbed distributions than
 2056 the test with the smaller sample size. This should be expected, as the detection of a small
 2057 proportion of odd events is generally difficult with a small sample size.
 2058



2059

2060 **Figure 45. OC curve for the normal distribution ($p=0$) compared to those for normal**
 2061 **distributions disturbed by an increasing proportion of exponentially ($\lambda=35\%$ LC)**
 2062 **disturbed values ($p = 5, 10, 15\%$). PTI tests using $n=10/30$ (left panel) and $n=24/72$**
 2063 **(right panel). (The reason for the missing left part of the OC curves is that the**
 2064 **coverage cannot become higher for these exponentially disturbed normal**
 2065 **distributions due to the presence of the heavy tail).**

2066

2067 3.4 Conclusion

2068 Based on the analysis presented above, it can be concluded that in situations with

2069

- 2070 • multi-modal distributions,
- 2071 • skewed distributions, and
- 2072 • heavy-tailed distributions,

2073

2074 the proposed PTI test provides similar or better control compared to situations with normally
 2075 distributed data having the same coverage. Thus, the risk for a low quality batch to be accepted
 2076 is smaller or equally low when data deviate from the normal distribution. The consumer is thus
 2077 well protected in these situations.

2078

2079 Further, it has been shown that for a high quality batch (*i.e.*, a batch with high coverage of the
 2080 target interval), the acceptance probability may be significantly reduced when data is non-
 2081 normally distributed. This provides producers with an incentive to avoid such situations.

2082

4 Simulation and Development of PTI Test Coefficients

4.1 Algorithm

This appendix details the algorithm used for determination of k_1 , k_2 , and f for the proposed PTI test. Before the algorithm can be used, the sample sizes for tier 1 and tier 2 must be specified. The minimum acceptable coverage level is set to 85% of the target interval $100 \pm 25\%$ LC.

The coefficients k_1 , k_2 and f are determined for a selected sample size (*e.g.* $n=10/30$) so that the probability of passing the test at the minimum acceptable level of coverage is 2.5% (for a mean on target) for the first tier and 5% for both tiers. The algorithm is accomplished in two iterations. First, initial estimates of k_1 and k_2 are determined using simulations (see section 4.3 below) assuming that the mean is 20% off target. This point was chosen so that the maximum sample standard deviation (MSSD) criterion would have no effect on the calculation of k_1 and k_2 (because the value of f is unknown at this stage). Using these preliminary estimates of k_1 and k_2 , f is estimated by simulations, assuming an on-target mean. (Recall that $MSSD=25f/k$). Once f has been determined, final values for k_1 and k_2 are calculated assuming an on-target mean and the determined f value. This procedure results in coefficients that provide the overall desired properties, *i.e.*, that iso-probability curves closely follow the corresponding iso-coverage curves, that the overall acceptance probability for the minimum acceptable coverage is 5%, and that the 1st tier acceptance probability is 2.5% when the mean is on target.

Here are the three steps required to complete the algorithm. For the given coverage (85%), n_1 and n_2 , and assuming a normal distribution:

- (1) Assuming a mean (μ) of 80% label claim and standard deviation (σ) corresponding to the given coverage, determine k_1 such that $\Pr(|100-m| + k_1s \leq 25) = 2.5\%$, where m and s denotes the mean and standard deviation for an independent sample of n_1 observations from the normal distribution $N(\mu, \sigma)$. Given k_1 , determine k_2 such that the overall pass rate for the test is 5% in this point.
- (2) Assuming an on-target mean ($\mu=100\%$ label claim) and standard deviation (σ) corresponding to the given coverage and using k_1 and k_2 calculated in step 1, determine f so that the overall pass rate of the test is 5%.
- (3) Assuming an on-target mean ($\mu=100\%$ label claim) and standard deviation (σ) corresponding to the given coverage and using f from step 2, determine k_1 such that $P(|100-m| + k_1s \leq 25 \text{ and } s \leq 25f/k_1) = 2.5\%$. Again, m and s denote the mean and standard deviation for an independent sample of n_1 observations from the normal distribution $N(\mu, \sigma)$. Using this estimate of k_1 , determine k_2 to obtain an overall pass rate of 5% in this point.

4.2 Example (n=10/30)

As an example, assume that $n_1=10$, $n_2=30$, and that the minimum acceptable coverage level is 85%.

- (1) If the true mean is at 80% of label claim, the standard deviation (σ) corresponding to 85% coverage can be calculated using the equation $\Pr(75 \leq x \leq 125) = 85\%$, where x is normally distributed with a mean of 80 and standard deviation σ . Under these conditions, $\sigma=4.82$.

Given a mean at 80% label claim and $\sigma=4.82$, k_1 and then k_2 can be determined using simulations. For this example, $k_1=2.25$ and $k_2=1.56$.

- (2) If the true mean is at 100% label claim and a coverage of 85% is assumed, the corresponding standard deviation is 17.4% LC. Given $\sigma=17.4$, $k_1=2.25$, and $k_2=1.56$ and assuming that the overall pass rate of the test at this point should be 5%, $f=0.839$. This value is determined by simulation.
- (3) If the true mean is at 100% label claim, $\sigma=17.4$, and $f=0.839$, k_1 and k_2 can be determined using simulation. For this example, $k_1=2.09$ and $k_2=1.59$.

4.3 Basis for Simulation

Simulation techniques have been used to determine the values of k_1 , k_2 , and f because it was not possible to determine all these values analytically for a multi-tiered test. Simulation techniques have also been used for determining OC curves and iso-probability curves throughout this report.

To illustrate the technique, consider the example of determining k_1 in step (1) of the example above. Given a mean of $\mu=80$ and a standard deviation of $\sigma=4.82$ and assuming that the data are normally distributed, it is desired to determine k_1 so that $\Pr(|100-m| + k_1s \leq 25) = 2.5\%$. One way to do this is to generate $n_1=10$ values from a normal distribution with the given mean and standard deviation and determine whether or not the criterion $|100-m| + k_1s \leq 25$ is passed for a given k_1 value. This can then be repeated for 100,000 samples drawn from the same distribution. The proportion of samples (out of the 100,000 samples) that pass this criterion is a good estimate of $\Pr(|100-m| + k_1s \leq 25)$. By repeating this for different values of k_1 , one can iterate to the k_1 value which gives $\Pr(|100-m| + k_1s \leq 25) = 2.5\%$.

For the example above, values of $k_1=1, 1.25, 1.5, 1.75, 2, 2.25, 2.5, 2.75, 3$ were initially used to determine the value of k_1 giving a result closest to the desired pass rate of 2.5%. It was determined that the desired value of k_1 was between 2.25 and 2.50. Next, values of $k_1=2.25, 2.26, 2.27, \dots, 2.50$ were used and the pass rate was determined. It was found that $k_1=2.25$ gave a pass rate closest to the desired level of 2.5%.

Values of k_2 and f were determined in a similar fashion. The criterion being evaluated for k_2 is $\Pr(\text{Pass overall test without MSSD criterion})$ for step (1) and $\Pr(\text{Pass overall test with MSSD criterion})$ for step (3). For determining f in step (2), the pass rate is evaluated for $\Pr(\text{Pass overall test with MSSD criterion})$.

4.4 SAS[®] Code

Included below is a SAS[®] code which can be used to determine k_1 , k_2 , and f for a two-tiered PTI test and a given sample size. This code has been validated by comparison with two other independently developed programs, which gave identical results. For a computer with 128 MB RAM, the simulations will take about 1½ hour. With a memory of 256 MB RAM, about ½ hour will suffice.

```
dm 'output; clear; log; clear;';
*****
*
*IPAC-RS_PTI_COEFFICIENTS ver 1.0 (DATE 2/2001)
*
*SAS(R) code to compute k1, k2 and f for two-tiered IPAC-RS proposed
*PTI test, providing 0.05 acceptance probability for 85% coverage of
*the 75-125% LC interval for normal distributed data (coverage determined
*by the set values for sd80 and sd100.
*
*User input:
*change n1=10 to desired n1 (1st tier sample size)
*change n2=30 to desired n2 (1st and 2nd tier total sample size) (n2>n1)
*change alpha1=0.025 to desired alpha1 (probability for 1st tier
*acceptance) alpha1<0.05
*change maxiter=100000 to desired maxiter (number of iterations)
*(note: a lower value for maxiter results in lower precision)
*****;

options pageno=1 linesize=80 pagesize=54 mprint;

*User input:
%let n1=10;          * Sample size in first tier;
%let n2=30;          * Sample size in both tiers combined;
%let alpha1=0.025;  * Chance to pass at tier 1;
%let maxiter=100000; * Number of iterations

*Constants
%let sd80=4.82;
%let mean80=80;
%let sd100=17.4;
%let mean100=100;

%macro k1;  * Macro for finding k1;

%macro search1(srchval);

*** First time through, get ball-park value for k1.  Second time, refine the
search;
data test;
  retain seed 1234396;
  alpha=&alpha1;
  sd80=&sd80;
  mean80=&mean80;

  array x {*} x1-x&n1;

  do tryk=&srchval;
    do iter= 1 to &maxiter;
```

```
sum=0;
sumsq=0;
do j=1 to &n1;
  x{j}=mean80 + sd80*rannor(seed);
  sum=sum+x{j};
  sumsq=sumsq+x{j}*x{j};
end;
sampmean=sum/&n1;
sampstd=sqrt((sumsq-&n1*sampmean*sampmean)/(&n1-1));
accvalue=abs(100-sampmean) + tryk*sampstd;

if accvalue le 25 then pass=1;
  else pass=0;
keep sampmean sampstd accvalue tryk iter pass alpha;
output;
end;
end;

run;

***proc print;
***  title 'test';
***run;

proc summary data=test mean n;
  var pass;
  by alpha tryk;
  output out=summary mean=pass;
run;

proc print;
  title 'Summary in search 1';
run;

*** Now find the one closest to alpha;
data summary;
  set summary;
  absdiff=abs(pass-alpha);
run;

*** Find 2 values closest to alpha;
proc sort;
  by absdiff;
run;

data findit;
  set summary(obs=1);
run;

proc print data=findit;
  title 'findit in search 1';
run;

%mend; *search1;

%let first=%str(1,1.25,1.5,1.75,2,2.25,2.5,2.75,3);

%search1(&first);

data findit;
  set findit;
```

```

    if pass<&alpha then call symput('goback','1');
        else call symput('goback','0');

    call symput('newsrch',put(tryk,4.2));
run;

%if &goback=1 %then %do;
    %search1(&newsrch-0.25 to &newsrch by 0.01);
%end;
%else %do;
    %search1(&newsrch to &newsrch+0.25 by 0.01);
%end;

data k1;
    set findit(keep=tryk rename=(tryk=k1));
run;

proc print data=k1;
    title 'k1';
run;

%mend; *k1;

%k1;

*****
*****
****                                     ****
*** Now that know what k1 is, try computing k2 ****
****                                     ****
*****
*****;

%macro k2; * Macro for finding k2;

*** Make a macro variable for k1 found in earlier search;
data k1;
    set k1;
    call symput('k1',put(k1,4.2));
run;

%macro search2(srchval);
*** First time through, get ball-park value for k2;

data test;
    retain seed 1234396;
    alpha=0.05;
    sd80=&sd80;
    mean80=&mean80;

    array x {*} x1-x&n2;

    do tryk=&srchval;
        do iter= 1 to &maxiter;
            sum=0;
            sumsq=0;

            sum1=0;
            sumsq1=0;

```

```

do j=1 to &n2;
  x{j}=mean80 + sd80*rannor(seed);
  sum=sum+x{j};
  sumsq=sumsq+x{j}*x{j};

  if j le &n1 then do;
    sum1=sum1+x{j};
    sumsq1=sumsq1+x{j}*x{j};
  end;
end;
sampmean=sum/&n2;
sampstd=sqrt((sumsq-&n2*sampmean*sampmean)/(&n2-1));
accval2=abs(100-sampmean) + tryk*sampstd;

sampm1=sum1/&n1;
sampstd1=sqrt((sumsq1-&n1*sampm1*sampm1)/(&n1-1));
accval1=abs(100-sampm1) + &k1*sampstd1;

if (accval1 le 25) or (accval2 le 25) then pass=1;
else pass=0;

keep sampmean sampstd sampm1 sampstd1 accval1 accval2 tryk iter
pass pass alpha;
output;
end;
run;

***proc print;
*** title 'test';
***run;

proc summary data=test mean n;
  var pass;
  by alpha tryk;
  output out=summary mean=pass;
run;

proc print;
  title 'Summary in search 2';
run;

*** Now find the one closest to alpha;
data summary;
  set summary;
  absdiff=abs(pass-alpha);
run;

*** Find 2 values closest to alpha;
proc sort;
  by absdiff;
run;

data findit;
  set summary(obs=1);
run;

proc print;

```

```

    title 'findit in search 2';
run;

%mend; *search2;

%let first=%str(1,1.25,1.5,1.75,2,2.25,2.5);

%search2(&first);

data findit;
  set findit;
  if pass<alpha then call symput('goback','1');
  else call symput('goback','0');

  call symput('newsrch',put(tryk,4.2));
run;

%if &goback=1 %then %do;
  %search2(&newsrch-0.25 to &newsrch by 0.01);
%end;
%else %do;
  %search2(&newsrch to &newsrch+0.25 by 0.01);
%end;

%mend; *k2;

%k2;

data ks;
  merge findit(keep=tryk rename=(tryk=k2))
    k1;
run;

proc print data=ks;
  title 'ks';
run;

*****
*****
****                               ****
**** Now that know k1 & k2, try computing MSSD ****
**** Use data that are distributed N(100,STD100) ****
****                               ****
*****
*****;

%macro MSSD; * Macro for finding MSSD;

*** Make a macro variable for k1 and k2 found in earlier search;
data ks;
  set ks;
  call symput('k1',put(k1,4.2));
  call symput('k2',put(k2,4.2));
run;

%macro search3(srchval);
*** First time through, get ball-park value for MSSD;

```

```
data test;
  retain seed 1234396;
  alpha=0.05;
  sd100=&sd100;
  mean100=&mean100;

  array x {*} x1-x&n2;

  do trymssd=&srchval;
    do iter= 1 to &maxiter;
      sum=0;
      sumsq=0;

      sum1=0;
      sumsq1=0;

      do j=1 to &n2;
        x{j}=mean100 + sd100*rannor(seed);
        sum=sum+x{j};
        sumsq=sumsq+x{j}*x{j};

        if j le &n1 then do;
          sum1=sum1+x{j};
          sumsq1=sumsq1+x{j}*x{j};
        end;
      end;
      sampmean=sum/&n2;
      sampstd=sqrt((sumsq-&n2*sampmean*sampmean)/(&n2-1));
      accval2=abs(100-sampmean) + &k2*sampstd;

      sampm1=sum1/&n1;
      sampstd1=sqrt((sumsq1-&n1*sampm1*sampm1)/(&n1-1));
      accval1=abs(100-sampm1) + &k1*sampstd1;

      if (accval1 le 25) and (sampstd1 le (25/&k1*trymssd)) then
passt1=1;
      else passt1=0;

      if passt1 or (accval2 le 25 and (sampstd le (25/&k2*trymssd))) then
pass=1;
      else pass=0;

      keep sampmean sampstd sampm1 sampstd1 accval1 accval2 trymssd iter
pass passt1 alpha;
      output;
    end;
  end;
run;

***proc print;
***  title 'test';
***run;

proc summary data=test mean n;
  var pass passt1;
  by alpha trymssd;
  output out=summary mean=pass passt1;
```

```
run;

proc print;
  title 'Summary in Search 3';
run;

*** Now find the one closest to alpha;
data summary;
  set summary;
  absdiff=abs(pass-alpha);
run;

proc sort;
  by absdiff;
run;

data findit;
  set summary(obs=1);
run;

proc print;
  title 'findit';
run;

%mend; *search3;

%let first=%str(0.75,0.775,0.80,0.825,0.85,0.875,0.90);

%search3(&first);

data findit;
  set findit;
  if pass<alpha then call symput('goback','0');
  else call symput('goback','1');

  call symput('newsrch',put(trymssd,4.3));
run;

%if &goback=1 %then %do;
  %search3(&newsrch-0.025 to &newsrch by 0.001);
%end;
%else %do;
  %search3(&newsrch to &newsrch+0.025 by 0.001);
%end;

%mend; *mssd;

%mssd;

data mssd;
  set findit(keep=trymssd passt1 rename=(trymssd=mssd));
run;

data allparms;
  merge mssd ks;
run;

proc print;
```

```

title 'Estimates of k1, k2 and MSSD before redo k1 and k2';
title2 "n1=&n1, n2=&n2, Alpha1=&alpha1, SD80=&sd80, SD100=&sd100";
var k1 k2 mssd passt1;
format passt1 5.3;
run;

*****
*****
****                                     ****
**** Redo estimate of k1, given MSSD      ****
****                                     ****
*****
*****;

*** Make a macro variable for mssd found in earlier search;
data allparms;
  set allparms;
  call symput('mssd',put(mssd,5.3));
run;

%macro kla; * Macro for finding k1 on 2nd time through;

%macro search1a(srchval);

*** First time through, get ball-park value for k1. Second time, refine the
search;
data test;
  retain seed 1234396;
  alpha=&alpha1;
  sd100=&sd100;
  mean100=&mean100;

  array x {*} x1-x&n1;

  do tryk=&srchval;
    do iter= 1 to &maxiter;
      sum=0;
      sumsq=0;
      do j=1 to &n1;
        x{j}=mean100 + sd100*rannor(seed);
        sum=sum+x{j};
        sumsq=sumsq+x{j}*x{j};
      end;
      sampmean=sum/&n1;
      sampstd=sqrt((sumsq-&n1*sampmean*sampmean)/(&n1-1));
      accvalue=abs(100-sampmean) + tryk*sampstd;

      if (accvalue le 25) and (sampstd le (25/tryk*&mssd)) then pass=1;
      else pass=0;
      keep sampmean sampstd accvalue tryk iter pass alpha;
      output;
    end;
  end;
run;

***proc print;
***  title 'test';

```

```
***run;

proc summary data=test mean n;
  var pass;
  by alpha tryk;
  output out=summary mean=pass;
run;

proc print;
  title 'Summary in redo of search 1';
run;

*** Now find the one closest to alpha;
data summary;
  set summary;
  absdiff=abs(pass-alpha);
run;

*** Find 2 values closest to alpha;
proc sort;
  by absdiff;
run;

data findit;
  set summary(obs=1);
run;

proc print data=findit;
  title 'findit in search 1';
run;

%mend; *search1a;

%let first=%str(1,1.25,1.5,1.75,2,2.25,2.5,2.75,3);

%search1a(&first);

data findit;
  set findit;
  if pass<&alpha then call symput('goback','1');
  else call symput('goback','0');

  call symput('newsrch',put(tryk,4.2));
run;

%if &goback=1 %then %do;
  %search1a(&newsrch-0.25 to &newsrch by 0.01);
%end;
%else %do;
  %search1a(&newsrch to &newsrch+0.25 by 0.01);
%end;

data k1;
  set findit(keep=tryk rename=(tryk=k1));
run;

proc print data=k1;
  title 'k1';
run;
```

```

%mend; *k1a;

%k1a;

*****
*****
****                               ****
**** Redo search for k2, given new k1 and MSSD ****
****                               ****
*****
*****;

%macro k2a; * Macro for finding k2;

*** Make a macro variable for k1 found in earlier search;
data k1;
  set k1;
  call symput('k1',put(k1,4.2));
run;

%macro search2a(srchval);
*** First time through, get ball-park value for k2;

data test;
  retain seed 1234396;
  alpha=0.05;
  sd100=&sd100;
  mean100=&mean100;

  array x {*} x1-x&n2;

  do tryk=&srchval;
    do iter= 1 to &maxiter;
      sum=0;
      sumsq=0;

      sum1=0;
      sumsq1=0;

      do j=1 to &n2;
        x{j}=mean100 + sd100*rannor(seed);
        sum=sum+x{j};
        sumsq=sumsq+x{j}*x{j};

        if j le &n1 then do;
          sum1=sum1+x{j};
          sumsq1=sumsq1+x{j}*x{j};
        end;
      end;
      sampmean=sum/&n2;
      sampstd=sqrt((sumsq-&n2*sampmean*sampmean)/(&n2-1));
      accval2=abs(100-sampmean) + tryk*sampstd;

      sampm1=sum1/&n1;
      sampstd1=sqrt((sumsq1-&n1*sampm1*sampm1)/(&n1-1));
      accval1=abs(100-sampm1) + &k1*sampstd1;

      if (accval1 le 25 and (sampstd1 le (25/&k1*&mssd))) or

```

```
(accval2 le 25 and (sampstd le (25/tryk*&mssd))) then pass=1;
else pass=0;

keep sampmean sampstd sampmnl sampstdl accval1 accval2 tryk iter
pass pass alpha;
output;
end;
end;
run;

***proc print;
*** title 'test';
***run;

proc summary data=test mean n;
var pass;
by alpha tryk;
output out=summary mean=pass;
run;

proc print;
title 'Summary in search 2';
run;

*** Now find the one closest to alpha;
data summary;
set summary;
absdiff=abs(pass-alpha);
run;

*** Find 2 values closest to alpha;
proc sort;
by absdiff;
run;

data findit;
set summary(obs=1);
run;

proc print;
title 'findit in search 2';
run;

%mend; *search2a;

%let first=%str(1,1.25,1.5,1.75,2,2.25,2.5);

%search2a(&first);

data findit;
set findit;
if pass<alpha then call symput('goback','1');
else call symput('goback','0');

call symput('newsrch',put(tryk,4.2));
run;

%if &goback=1 %then %do;
%search2a(&newsrch-0.25 to &newsrch by 0.01);
```

```
%end;
%else %do;
    %search2a(&newsrch to &newsrch+0.25 by 0.01);
%end;

%mend; *k2a;

%k2a;

data ks;
    merge findit(keep=tryk rename=(tryk=k2))
           k1;
    mssd=input(symget('mssd'),5.3);
run;

proc print data=ks;
    title 'Final Estimates of k1, k2 and MSSD';
    title2 "n1=&n1, n2=&n2, Alpha1=&alpha1, SD80=&sd80, SD100=&sd100";
    var k1 k2 mssd;
    format mssd 5.3 k1 k2 4.2;
run;
```

2852

2853

5 Glossary of Abbreviations, Symbols and Terms

2854

AV	Acceptance Value: $AV = 100-m + ks$.
DCU	Dose Content Uniformity.
DDU	Delivered Dose Uniformity.
f	Factor used to calculate MSSD: $MSSD = 25f/k$.
FDA DCU test	The <i>Dose Content Uniformity</i> tests described in Sections III.F.1.i and III.F.2.h of the <i>Metered Dose Inhaler and Dry Powder Inhaler Drug Products CMC Draft Guidance</i> . An identical test for <i>Spray Content Uniformity</i> is described in Sections III.F.1.g and III.F.2.p of the <i>Nasal Spray and Inhalation Solution, Suspension, and Spray Drug Products CMC Draft Guidance</i> .
FDA TCL test	The <i>Dose Content Uniformity Through Container Life</i> test, described in Sections III.F.1.j and III.F.2.i of the <i>Metered Dose Inhaler and Dry Powder Inhaler Drug Products CMC Draft Guidance</i> . The corresponding test for sprays described in Sections III.F.1.h and III.F.2.q of the <i>Nasal Spray and Inhalation Solution, Suspension, and Spray Drug Products CMC Draft Guidance</i> is slightly different.
FDA DCU&TCL test	Combined application of both the FDA DCU and the FDA TCL test, see Appendix 1, Section 1.3 for details.
k_1, k_2	Acceptability coefficients used in calculating the Acceptance Value for 1 st and 2 nd tier.
L	Limit for Acceptance Value.
LC	Label claim (<i>i.e.</i> the target dose), here, delivered dose label claim.
m	Overall Sample mean (arithmetic average).
m_{LS}	Life-stage sample mean.
μ	Population mean (<i>i.e.</i> , true mean of a batch).
MSSD	Maximum Sample Standard Deviation: $MSSD = 25f/k$.
NMT	Not more than.
n_1	Sample size in the 1 st tier.
n_2	Total sample size for both tiers.
OC curve	Operating Characteristic curve.
OINDP	Orally Inhaled and Nasal Drug Products.
PTI test	Parametric Tolerance Interval test.
s	Sample standard deviation.
σ	Population standard deviation (<i>i.e.</i> , true standard deviation of a batch).
SD	Standard deviation.
T	Target.
n=10/30	$n_1=10, n_2=30$.
n=24/72	$n_1=24, n_2=72$.

2855

2856

2856

Acceptance rate	Same as Probability to accept.
Consumer protection	Preventive measures which protect the consumer from batches of product at or below the limiting quality.
Consumer risk	Probability of accepting a batch at the limiting quality (here, 5%).
Container	That which contains the medicinal formulation (e.g., an MDI, DPI, or a single-dose blister or capsule).
Coverage	Proportion of the population (batch) that falls within the specified target interval. For a normal distribution with mean μ and standard deviation σ , the coverage of the 75-125% LC target interval may be calculated from: $\text{Coverage}_{75-125} = \frac{1}{\sigma\sqrt{2\pi}} \int_{75}^{125} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx$
Density	In statistics, a function $f(x)$ describing the shape of a particular distribution. For a normal distribution with mean μ and standard deviation σ , the density is $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right].$
Device	That which is used to administer the medicinal formulation.
Dose	Amount of drug delivered after actuating the inhaler the minimum number of times specified on the label.
Exponential distribution	An exponential distribution with parameter λ is defined by the density $f(x) = \exp(-x/\lambda)/\lambda$, $x \geq 0$, $\lambda > 0$.
Gamma function	The gamma function $\Gamma(n)$ is defined by the integral $\Gamma(n) = \int_0^{\infty} t^{n-1} e^{-t} dt.$ <p>When n is a positive integer, $\Gamma(n) = (n-1)!$.</p>
Inhaler	A combination of the container in which the medicinal formulation is packaged and the device that dispenses it.
Inner limits	For FDA tests, the 100±20% LC interval.
Iso-coverage curve	In the coordinates σ -vs- μ , a curve passing through those pairs of (μ, σ) that correspond to the same coverage of a given target interval.
Iso-probability curve	In the coordinates σ -vs- μ , a curve passing through those pairs of (μ, σ) that corresponds to the same probability of passing a test.
Limiting quality	Batch quality such that a sample has a low (here, 5%) probability of passing the test, here 85% coverage of the target interval 100±25% LC.
Non-parametric test	A test that does not assume data to follow any particular distribution (e.g., a "counting" test that counts the number of observation within a certain fixed range).
Monotonic	A multi-dose product is monotonic if the level of the middle life-stage is typically within the range formed by the levels of the beginning and end life stages

Operating Characteristic curve	A plot of the probability to pass a test as a function of a quality measure for the batch (<i>e.g.</i> , standard deviation, coverage).
Outer limits	For FDA tests, the 100±25% LC interval.
Parametric test	A test that assumes data to follow a particular distribution (<i>e.g.</i> , normal), which depends on one or more parameters (<i>e.g.</i> , mean, standard deviation) of the distribution.
Probability to accept (Acceptance probability)	Probability that a sample randomly drawn from the batch meets the acceptance criteria of a test.
Producer risk	Probability of rejecting a batch of a quality that exceeds the limiting quality.
Rejection probability	Rejection probability = 100% - Acceptance probability.
Sample	A finite set of data collected from the population.
Sample size	The number of data points (observations) used in a test.
Sampling plan	Rules describing how a sample is collected.
Stratified sampling plan	A sampling plan in which objects of different sub-classes are randomly sampled in pre-determined proportions.
Target	100% LC.
Target interval	<i>Here</i> , the interval 75-125% of the label claim.
Zero tolerance requirement	A requirement that no value outside a pre-defined limit is allowed.

2857