Gaps in statistical approaches to control of delivered dose uniformity throughout product lifecycle

Quality management should be adjusted throughout product lifecycle to minimize the probability of false rejection and false acceptance errors.

Greg Larner, Dave Christopher, Helen Strickland, Emil Friedman, David Starbuck, Michael Golden and Svetlana Lyapustina

International Pharmaceutical Aerosol Consortium on Regulation and Science (IPAC-RS) Delivered Dose Uniformity Subgroup

Abstract

This article discusses gaps and challenges in current approaches to control of delivered dose uniformity of orally inhaled products. A simulated data example is presented to illustrate the problem. A lasting solution would be the adoption of sample-independent quality standards and statistical process controls, for example, by incorporating in the pharmaceutical industry some of the statistical methods standardized by ASTM, ISO and other standard-setting organizations, appropriately adjusted throughout a product's lifecycle (from early to late development to commercial production).

Introduction

Delivered dose uniformity (DDU) is one of the critical quality attributes of inhalation products. The currently recommended regulatory test for batch release is a parametric tolerance interval two-one-sided test (PTI-TOST) proposed by the US Food and Drug Administration (FDA) at an Advisory Committee meeting in 2005.[1] The formal FDA guidance (still in draft form, dated 1998) includes a "counting" (non-parametric) test for DDU.[2] Both of these tests are demonstration tests, meaning that they use a sample to make a decision about batch disposition (pass/fail). These tests are also transactional, meaning that in judging quality of a given batch, only information from that batch's sample is taken into consideration. Furthermore, the counting test does not contain any explicit statement as to the required properties of the batch as a whole (e.g., permissible variation of the batch mean, or allowable within-batch standard deviation of the dose). The PTI-TOST does specify the confidence and coverage with respect to batch release criteria (the test demonstrates with 95% confidence that no more than 6.25% of the dose distribution lies outside either side of the target interval of 80-120% Label Claim), although the stringency of these requirements may or may not be appropriate for the majority of inhaled products.

In parallel, the US Pharmacopeia (USP) sets forth the requirements for DDU of aerosol products.[3] The USP General Notices and Requirements[4] clarify that pharmacopeial tests apply only to the units tested, and are not intended for making judgment about any batch, but that the product units must meet the USP requirements "at all times in the life of the article from production to expiration."[5] Re-

cently, the FDA also acknowledged that the USP criteria are insufficient for batch release.[6] Setting aside the question of whether these USP standards do or do not, or even if they should, match the FDA requirements, the point to highlight for the purposes of this article is that the USP does not currently aim to set any explicit standards for the batch as a whole. Nevertheless, USP testing is historically used by industry to demonstrate compliance "with pharmacopeial quality standards" (which, according to the USP, means compliance of just the tested units with just that USP test). Similar to the FDA DDU tests, the USP's is a demonstration test, and similar to the FDA 1998 guidance test, the USP's is a "counting" (non-parametric) test.

Requirements for the batch as a whole could be back-calculated from any of the tests described above, but the solution for such an inferred "batch standard" is not unique, and will depend on the interpretation of the way a particular test is to be implemented, as well as on a series of procedural and distributional assumptions. In other words, a given demonstration test may correspond to a family of "batch standards."

Overall, the regulatory and pharmacopeial approaches are not well aligned with each other, creating challenges discussed elsewhere,[7] which in principle apply to all pharmaceuticals, not only orally inhaled products (OIPs). For the purposes of this article, the focus is on the difficulty that is unique to inhalation drugs: a relatively high false rejection rate in DDU tests, as experienced by a significant number of companies manufacturing OIPs.[8-10] The statistical term "false rejection rate" refers to the probability (or frequency) of rejecting a batch of acceptable quality, when quality is understood as a sample-independent property of the batch. The rejection occurs through failing to meet the requirements of a demonstration test on a sample. For OIPs, the root cause of this problem could be traced to the combination of technological challenges and very stringent regulatory requirements. Aerosol doses are more variable than doses in solid tablets even when the composition of the formulation mix going into inhalers is tightly controlled, the manufacturing technology is top-of-the-line, and design of OIP delivery devices is state-of-the-art. Graphically, for OIPs, the area of the product's natural variation (e.g., on a plot of batch_mean-vs-batch_standard_deviation) often takes up the entire span of the "acceptance region" calculated from the regulatory specifications. This means that batches that are within the acceptance region could fail the sample demonstration test by mere chance if they come close to the border of the "acceptance region." Unfortunately for OIPs, the border of the acceptance region is often set close to the operational region. This effect was discussed at length previously[11] and will be briefly illustrated below.

An additional drawback of the current approach to DDU control is that demonstration tests only use information from a single batch (i.e., the batch under consideration) and do not take into account the previously manufactured batches. (As a reminder, the specifications themselves are set based on a limited number of batches in late development, and specifications of the FDA and USP tests are pre-determined, not based on any batches' data). This practice fails to make use of the valuable information about the state of the manufacturing process as a whole and the comparison of the current batch to the typical batches over time. A strict reliance on a transactional approach prevents the possibility of combining trending information and batch acceptance control criteria as one of the options for control strategy, which has been receiving more attention in recent public conferences.[6,12] Furthermore, demonstration tests in the pharmaceutical industry use fixed sample size and do not allow the manufacturer to expand or reduce the size of the sample based on the current information obtainable from the manufacturing line. A more comprehensive approach would be to learn from other industries and establish sample-independent quality standards (or mechanisms for establishing such standards for a given product based on development data), adopt methods in the pharmaceutical industry for taking into account data from previously manufactured batches (e.g., using ASTM and other existing methodologies as appropriate) and introduce switching rules to collect more or less information as justified by the data from the ongoing manufacturing process and preceding test results (again, using methodologies developed and standardized through ASTM International [formerly the American Society for Testing and Materials] and the ISO [International Organization for Standardization]).

A data simulation

To illustrate the problems mentioned in the Introduction, DDU data was simulated for 25 lots (batches), using the following parameters that would not be unusual for an approved inhaled product (e.g., as can be seen from the IPAC-RS DDU Database[13]):

- Batch Overall Mean = 98.0% Label Claim
- Within Batch Pooled Standard Deviation = 6.00
- Between Batch Standard Deviation = 4.00

The first 5 lots were used for "validation," with sample size n=50 inhalers tested per batch. The next 20 lots were "commercial," with sample size n=10 inhalers tested per batch. In this simulation, one dose was collected from each inhaler. The results are presented in Figure 1. (Supplemental data for Figure 1 is presented in Table 1.) All doses lie within the target interval recommended by the FDA PTI-TOST with the target interval of 80-120% Label Claim for delivered dose. All commercial batches in this example pass the USP test,[3] as well as the FDA 1998 counting test. Yet 15 of these batches would have failed the FDA PTI-TOST tier 1 criteria [target interval 80-120%; maximum allowed area under each tail 6.25% of all doses]. Of these, 5 batches would have ultimately failed tier 2 of the FDA PTI-TOST if the sample mean and standard deviation were the same as in tier 1. This example shows that the same data could lead to different answers depending on which of the "standards" (USP <601>, FDA counting 1998 test or FDA PTI-TOST) is used to assess the data. The underlying cause of this discrepancy is that these tests have different quality standards at the same confidence level, or the same quality standard at different confidence levels, or a combination of the two causes. The lack of a quality standard makes it difficult to effectively manage product quality. A central theme of this article is that one needs to start by defining a suitable quality standard derived from a quality target product profile (as defined in the International Conference on Harmonisation ICH Q8R2 document).[14]

Moreover, it is important that the standard be suitable for the products at hand because, as can be seen in this example, a product with not unusual characteristics can face a relatively high rejection rate with PTI-TOST. Such stringency not only wastes resources associated with the production, testing and investigation of these (possibly inappropriately) rejected batches, but they may also be masking the real signal of a process change or quality change, due to the "noise" created by these signals. Another downside of inappropriate rejection is that a product suitable for use could possibly not reach the patient.

Setting a wider target interval, e.g., at 75-125% Label Claim, could have been more appropriate for this and similar products. By using operating characteristic curves, such changes in test requirements could be introduced without increasing the probability of false acceptance rates.[15] Currently, however, there is no pathway for justifying wider limits, even though the target interval of 80-120% Label Claim is itself not explicitly justified, but rather is

Figure 1

Illustration of the concept of false rejection. Results for the example discussed in the text. Horizontal lines represent the target interval recommended by FDA PTI-TOST (80-120% Label Claim for dose). N=50 for the first 5 lots; N=10 for the next 20 lots. All lots came from the same manufacturing process with Batch Overall Mean = 98.0% Label Claim; Within Batch Pooled Standard Deviation = 6.00; Between Batch Standard Deviation = 4.00. The line tracks the sample mean.

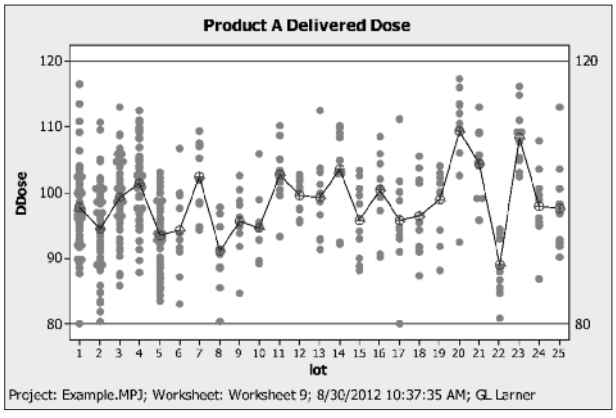


Table 1

Supplemental data for Figure 1.

One-way ANOVA: DDose versus lot

Source	DF	SS	MS	F	P
lot	24	7553.2	314.7	8.75	0.000
Error	425	15280.9	36.0		
Total	449	22834.1			

S = 5.996 R-Sq = 33.08% R-Sq(adj) = 29.30%

Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev
1	50	97.70	6.67
2	50	94.54	6.79
3	50	99.27	5.79
4	50	101.39	5.97
5	50	93.54	5.06
6	10	94.32	6.84
7	10	102.39	5.60
8	10	91.23	5.73
9	10	95.66	5.02
10	10	94.65	5.04
11	10	102.70	4.74
12	10	99.65	2.79
13	10	99.20	6.36
14	10	103.56	6.58
15	10	95.86	5.77
16	10	100.49	6.28
17	10	95.85	8.07
18	10	96.40	6.24
19	10	98.88	5.55
20	10	109.34	7.39
21	10	104.43	5.25
22	10	88.97	4.72
23	10	108.35	4.67
24	10	97.94	5.66
25	10	97.64	6.83

91.0 98.0 105.0 112.0

Pooled StDev = 6.00

pre-determined by the FDA proposal which focuses on the demonstration test. Neither the FDA nor USP nor any other standard-setting body currently explains how DDU specifications (including the width of the target interval) for the final product could be derived from the data obtained in early through late development. This lack of statistical procedures for transitioning from early development to the final product with respect to DDU controls, and particularly the acceptance criteria, represents one of the key gaps in the current approaches. By contrast, for demonstration tests for other product attributes, data available from development batches are typically used to set regulatory acceptance criteria.

Another aspect illustrating the need to look at the bigger picture is presented in Figure 2. The same data as in Figure 1 is used here to plot sample standard deviations for each lot. This so-called S-Chart indicates that for n=10, sample standard deviations should not exceed 10.016 when the process is stable and typical of the previously manufactured ("historical") batches. This limit for sample standard deviation could be adjusted for increased sample sizes, which may be needed in validation or other circumstances. This alternative method of monitoring and controlling quality will reduce the rate of false rejections. But adopting such a strategy requires a wider focus, not limited to the batch in question but taking into account the manufacturing process as a whole, and for the particular batch in the context of that overall process. The narrow focus on one batch at a time represents a second key gap in the current approaches to DDU control.

Conceptually, the FDA does recognize the need for changing testing requirements throughout the product's lifecycle. The FDA thinking is illustrated in Figure 3, which is based on the FDA 2011 Process Validation guidance.[16] This guidance and subsequent presentations by FDA scientists at conferences[17] suggest that in earlier stages of product development, when there is less understanding of the product and manufacturing process, the sampling and testing from each individual batch should be more extensive. As more information and process understanding are gained and relevant in-process controls are established, a "stepped-down" sampling from the batch could be adopted with justification. The guidance is silent, however, on the types of information that constitute appropriate justification. This lack of information about the transition from extensive to stepped-down sampling is the third key gap in current approaches to DDU control.

Figure 2

Sample standard deviations from the lots presented in Figure 1.

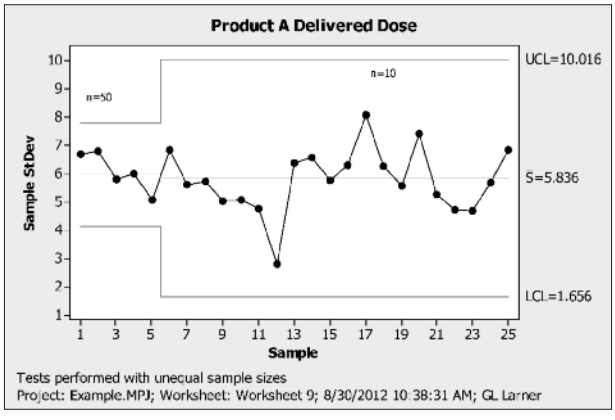
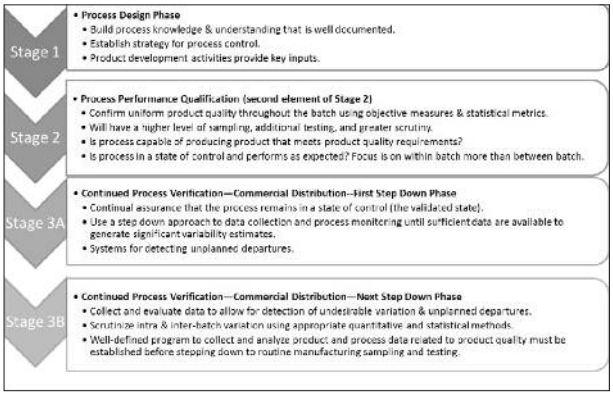


Figure 3

Lifecycle phases based on the FDA 2011 Process Validation guidance.



Industry, regulators and other stakeholders should work together to bridge the key gaps identified above, so that the quality control system could become more sound and decision making more contiguous, with information and knowledge from previous stages of a product's lifecycle becoming integrated into subsequent test requirements, creating a balance of in-process and finished-product testing. Such a system is necessary because the risks and purpose of testing change throughout the product's lifecycle, and the basis for making decisions regarding quality grows and evolves during development. For example, in the earlier phases, for material to be used in controlled clinical trials, wider acceptance ranges may be appropriate (and may even be necessary at this early stage) because the risks being managed are different than those of commercial product, which is used in a wider population without the close monitoring used in clinical trials. In later product lifecycle phases, the rela-

tive importance of maintaining a consistent manufacturing process rises, ensuring that product is consistent with material shown acceptable in clinical trials. The focus in commercial production shifts to distinguishing DDU results that are typical and expected from those that are atypical or unexplained and unexpected.

Currently missing is a statistically described implementation plan for the progression of quality controls from testing early-development products to testing commercial lots. To put the FDA recommendations into practice, a system must be established that clearly addresses the following three crucial areas, which are often confused when left undefined:

Quality Standard—Explicit requirements for characteristics of the entire population of doses for a given product or a group of similar products, as part of chemistry, manufacturing and controls (CMC) requirements. Statistically speaking, a quality standard could specify such parameters as the proportion of dose distribution within the target interval (so-called "coverage"). Note that a quality standard in itself does not involve sampling, parameter estimation or degree of uncertainty/risk (i.e., confidence). The definition of the "population" of doses, however, should be specified when a quality standard is set. For example, it may refer to a batch of finished product, or all past and future batches of a given product from a given manufacturer, or all batches of a given product from all manufacturers, etc. All of the above scenarios could be appropriate, but whatever the choice, it should be explicitly stated. Furthermore, the quality standard should be appropriate for the product(s) for which it is intended, i.e., minimizing both false acceptance and false rejection probabilities while using the goal posts (target intervals) driven by the user's (patient's) clinical requirements. A public quality standard should integrate input from chemists, clinicians and statisticians, perhaps through a process managed by a standard-setting body such as the USP. Until such public standards are established and justified as appropriate for particular products, sponsors could propose quality standards for their products to the FDA based on their product's data (both clinical and CMC).

Quality Decision Framework—A structured approach for determining whether the quality standard has been met. (In statistical terms, this is a probability assessment.) The quality decision framework does involve sampling to estimate dose population parameters with specified uncertainty (or confidence). The objective of decision making should be clearly understood because it may affect the level of acceptable uncertainty (or confidence) and/or choice of demonstration test. As discussed above, in earlier phases of development, lower confidence levels could be appropriate but in commercial production higher confidence would be required. The quality decision framework must be built on sound statistical principles, e.g., as learned from experience of other industries as well as relevant standards established by ISO,[18] ASTM,[19] and ASQ(C) [formerly the American Society for Quality Control].[20]

Demonstration Test—A statistical tool that enables decision making in the quality decision framework (i.e., a mechanism for quantifying probability). Acceptance limits for decision making must align with the objective of testing at a given stage in the product's lifecycle and with the required level of confidence.

For the demonstration test, the current (traditional) approach in the pharmaceutical industry is transactional testing, meaning that each batch is judged based only on the performance of the sampled units from that batch. As illustrated earlier in this article, this type of demonstration test aggravates the problem of stringent FDA tests for DDU of OIPs. By contrast, demonstration testing would be less likely to falsely reject batches of suitable quality if it became routine practice in the pharmaceutical industry to:

(i) combine multiple acceptance plans and approaches, and use pre-determined switching rules (e.g., by using the concepts and principles in existing standards, such as ASTM-2810 and ASTM-2709);

(ii) apply, and formally integrate into the quality decision framework, statistical process control for monitoring and controlling DDU quality (possibly using multi-variate statistical process controls during the manufacturing process); and

(iii) aggregate and use historical data to better estimate variability of the overall process and to better assess the typical or atypical nature of the batch at hand. More concrete examples and potential implementation proposals are still under development and will be discussed in future articles and conference presentations.

Conclusions

To maximize efficient use of information and resources, the approaches to quality management should be adjusted throughout product lifecycle. The ability to switch to alternative demonstration tests and acceptance plans in a pre-defined manner and with solid justification can help minimize the

probability of statistical false rejection of suitable-quality product while maintaining an acceptably low probability of statistical false acceptance of unsuitable-quality product. Reduction of both error types benefits the consumer. This flexible approach will enable the focus on assessing product quality rather than on testing quality into the product. These approaches would be beneficial to all pharmaceutical dosage forms, but for OIPs the problem is most pressing due to the combination of a variable nature of aerosols and stringent regulatory specifications. For manufacturing in the 21st century, the balance of control should shift emphasis from the finished product to in-process testing, and comparison with historical data should become part of the quality paradigm as a product progresses from early development to commercial production.

References

1. M.M. Nasr. Parametric Tolerance Interval Test for Delivered Dose Uniformity (DDU) Working Group Update (2005). http://www.fda.gov/ohrms/dockets/ac/05/slides/2005-4187S1_13_Nasr.ppt

2. FDA, Center for Drug Evaluation and Research (CDER). Draft Guidance for Industry. Metered Dose Inhaler (MDI) and Dry Powder Inhaler (DPI) Drug Products - Chemistry, Manufacturing, and Controls Documentation (1998). http://www.fda.gov/downloads/Drugs/Guidances/ucm070573.pdf

3. USP. Pharmacopeial Forum 39(1): General Chapter <601> Aerosols, Nasal Sprays, Metered-Dose Inhalers, and Dry Powder Inhalers. http://www.usp.org/usp-nf/pharmacopeial-forum

4. USP. General Notices and Requirements. http://www.usp.org/sites/default/files/usp_pdf/EN/USPNF/revisions/2013-07-09_general_notices_usp37-nf32_final.pdf

5. USP. PharmForum 38(6) Stimuli to the Revision Process: Methods for Measuring Uniformity in USP. http://www.usp.org/usp-nf/pharmacopeial-forum

6. FDA. Questions and Answers on Current Good Manufacturing Practices, Good Guidance Practices, Level 2 Guidance - Production and Process Controls (2013). http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm124782.htm#15

7. S. Lyapustina. A Test is Not a Quality Standard: IPAC-RS Response to the Pharmacopeial Forum Stimulus Article "Methods for Measuring Uniformity in USP." Respiratory Drug Delivery Europe 2013 (2013), Vol 2, pp 309-312.

8. ITFG/IPAC. Initial Assessment of the ITFG/IPAC Dose Content Uniformity Database by the CMC Specifications Technical Team of the ITFG/IPAC Collaboration (2000). http://ipacrs.com/PDFs/Initial_Assessment_of_Dose.PDF

9. B. Olsson, S. Lyapustina. Debating the Parametric Operating Curves for DDU Tests on Marketed Inhalers. Respiratory Drug Delivery IX (2004), Vol 1, pp 135-142.

10. B. Olsson, D. Sandell. Delivered Dose Uniformity Testing: IPAC-RS Advocacy and Justification. Respiratory Drug Delivery VIII (2002), Vol 1, pp 115-122.

11. IPAC-RS. A Parametric Tolerance Interval Test for Improved Control of Delivered Dose Uniformity of Orally Inhaled and Nasal Drug Products (2001). See section 7.2. http://ipacrs.com/PDFs/IPAC-RS_DDU_Proposal.PDF

12. ISPE Annual Meeting, Session "Stratified Sampling Plans." http://www.ispe.org/2013-annual-meeting/program-schedule

13. Slide 11 in B.Olsson. Debating the Operating Curves for DDU Tests on Marketed Inhalers. Presented at RDD 2004. http://ipacrs.com/PDFs/RDD_IX_DDU.PPT

14. ICH Q8R2. Pharmaceutical Development (2009). http://www.ich.org/products/guidelines/quality/article/quality-guidelines.html

15. S. Novick, D. Christopher, M. Dey, S. Lyapustina, M. Golden, S. Leiner, B. Wyka, H-J. Delzeit, C. Novak and G. Larner. A Two One-Sided Parametric Tolerance Interval Test For Control of Delivered Dose Uniformity* – Part 2 – Effect of Changing Parameters. AAPS PharmSciTech 2009 DOI: 10.1208/s12249-009-9269-3. Vol 10. No 3, September 2009. pp. 841-849. http://www.pharmagateway.net/ArticlePage.aspx?DOI=10.1208/s12249-009-9269-3

16. FDA. Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), Center for Veterinary Medicine (CVM)/Guidance for Industry. Process Validation: General Principles and Practices (2011). http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm070336.pdf

17. Richard Friedman. Introduction: SPC as a Tool to Monitor the State of Control. Presentation at the IFPAC Annual Meeting (Baltimore, MD) on January 22-25, 2013.

18. International Organization for Standardization. http://www.iso.org/iso/home.html

19. Formerly the American Society for Testing and Materials, now ASTM International. http://www.astm.org/

20. Formerly the American Society for Quality and the American Society for Quality Control (ASQC), now ASQ. http://asq.org/index.aspx

Authors and Affiliations

Greg Larner (Manager, Scientific and Laboratory Services, Pfizer)

Dave Christopher (Head, Nonclinical and Pharmaceutical Sciences Statistics, Merck Research Laboratories)

Helen Strickland (Statistical Consultant, GSK)

Emil Friedman (Principal Scientist/Statistician, Pharmaceutical Analytics, MannKind Corporation)

David Starbuck (Scientist, GSK)

Michael Golden (Senior Vice President, Regulatory Affairs and Quality, Pearl Therapeutics)

Svetlana Lyapustina (Senior Science Advisor, Drinker Biddle & Reath, LLP)

Corresponding author: Svetlana Lyapustina, Senior Science Advisor, Drinker Biddle & Reath, LLP, 1500 K Street NW, Washington, DC 20005, US, Tel: +1 202 230-5179, svetlana.lyapustina@dbr.com. Website: www.ipacrs.com.