

Statistical Methods for Assessment of Blend Homogeneity

Camilla Madsen

**LYNGBY 2002
IMM-PHD-2002-99
ATV Erhvervsforskerprojekt EF 767**

IMM

IMM-PHD-2002-99
ATV Erhvervsforskerprojekt EF 767

ISSN: 0909-3192
ISBN: 87-88306-15-1

© Copyright 2002 by Camilla Madsen.

This document was prepared with L^AT_EX and printed by DTU-tryk.

Preface

This thesis has been prepared at the department Informatics and Mathematical Modelling (IMM), Technical University of Denmark (DTU), and at the pharmaceutical company Novo Nordisk A/S in partial fulfilment of the requirements for the industrial Ph.D. degree within the Mathematical Ph.D. Program at DTU.

The thesis is concerned with statistical methods to assess blend uniformity in tablet production.

Lyngby, May 2002.

Camilla Madsen

Acknowledgements

First of all, I wish to thank my academic supervisor, Prof. Poul Thyregod, IMM, DTU for many hours of discussions and support. His guidance has been crucial for the outcome of this project and for that I am very grateful. I also want to thank my three industrial supervisors at Novo Nordisk A/S, Jørgen Iwersen, Quality Support Statistics, Per Grønlund, SDF Pilot Plant and Charlotte Tvermoes Rezai, SDF Production, for their help and for our enlightening discussions throughout the project. Further I would like to thank Jørgen Iwersen for taking the initiative to this project and for making an effort on finding the right collaborators, and Per Grønlund and Charlotte Tvermoes Rezai for their help and willingness to always answer my questions regarding pharmaceutical matters. They also deserve thanks for their effort in trying to find the time for experiments in a very tight production schedule.

I also would like to thank colleagues in the Quality Support Statistics department, Novo Nordisk A/S and at IMM, DTU for a good and pleasant scientific and social environment. Águsta Hafliðadóttir as well as my two officemates Dorte Rehm and Dorte Vistisen deserves special acknowledgements for many pleasant and constructive hours.

Last but not least I am grateful to my fiancée, my family and my friends for their support, patience and encouragement during the hard parts of this work.

Summary

In this thesis the use of various statistical methods to address some of the problems related to assessment of the homogeneity of powder blends in tablet production is discussed.

It is not straight forward to assess the homogeneity of a powder blend. The reason is partly that in bulk materials as powder blends there is no natural unit or amount to define a sample from the blend, and partly that current technology does not provide a method of universally collecting small representative samples from large static powder beds.

In the thesis a number of methods to assess (in)homogeneity are presented. Some methods have a focus on exploratory analysis where the aim is to investigate the spatial distribution of drug content in the batch. Other methods presented focus on describing the overall (total) (in)homogeneity of the blend. The overall (in)homogeneity of the blend is relevant as it is closely related to the (in)homogeneity of the tablets and therefore critical for the end users of the product.

Methods to evaluate external factors, that may have an influence on the content in blend samples, as e.g. sampling device, have been presented. However, the content in samples is also affected of internal factors to the blend e.g. the particle size distribution. The relation between particle size distribution and the variation in drug content in blend and tablet samples is discussed.

A central problem is to develop acceptance criteria for blends and tablet batches

to decide whether the blend or batch is sufficiently homogeneous (uniform) to meet the need of the end users. Such criteria are most often criteria regarding sample values rather than criteria for the quality (homogeneity) of the blend or tablet batch. This inherently leads to uncertainty regarding the true quality of a specific blend or batch. In the thesis it is shown how to link sampling result and acceptance criteria to the actual quality (homogeneity) of the blend or tablet batch. Also it is discussed how the assurance related to a specific acceptance criteria can be obtained from the corresponding OC-curve.

Further, it is shown how to set up parametric acceptance criteria for the batch that gives a high confidence that future samples with a probability larger than a specified value will pass the USP three-class criteria.

Properties and robustness of proposed changes to the USP test for content uniformity are investigated by the use of simulations, and single sampling acceptance plans for inspection by variables that aim at matching the USP proposal have been suggested.

Resumé (in Danish)

Denne afhandling omhandler brugen af statistiske metoder til at belyse forskellige problemstillinger i forbindelse med vurdering af homogeniteten af en pulverblanding i tabletfremstilling.

Det at bestemme homogeniteten af en pulverblanding er ikke simpelt. Det skyldes dels at bulkmaterialer som pulverblandinger ikke indeholder en naturlig enhed eller mængde, der kan afgrænse en prøve fra blandingen og dels at der med den nuværende teknologi ikke findes en universel metode til at indsamle små repræsentative prøver fra store statiske pulverblandinger.

I afhandlingen er forskellige metoder til at vurdere homogenitet beskrevet. De første metoder kan anvendes i forbindelse med explorative undersøgelser, hvor formålet er at undersøge fordelingen af aktivt stof i blandingen. De sidste metoder har til formål at beskrive den overordnede (totale) homogenitet i blandingen. Den overordnede homogenitet i blandingen er relevant, da den har betydning for homogeniteten af tabletterne, og derfor er den kritisk for de endelige forbrugere af tabletterne.

Metoder til at vurdere ydre faktors betydning for indholdet af aktivt stof i prøver fra blandingen er blevet diskuteret. En ydre faktor kan f.eks. være det redskab, prøverne udtages med. Indholdet af aktivt stof i prøver fra blandingen samt i tabletterne afhænger også af indre faktorer som f.eks. partikel størrelsesfordelingen. Sammenhængen mellem partikel størrelsesfordeling og variation i indholdet af aktivt stof i prøver fra blandingen og tabletbatchen er blevet diskuteret.

Et central problem er at opstille accept kriterier for blandinger og tablet batche, der sikrer at homogeniteten er tilfredsstillende i forhold til de endelige forbrugeres behov. Sådanne accept kriterier bliver ofte formuleret som krav til resultatet af stikprøven i stedet for mere direkte som krav til kvaliteten (homogeniteten) af blandingen eller tabletbatchen. Sådanne krav fører nødvendigvis til usikkerhed angående den sande kvalitet af den enkelte blanding eller tabletbatch. I afhandlingen er sammenhængen mellem på den ene side stikprøveresultat og accept kriterium og på den anden side kvaliteten (homogeniteten) af en blanding eller tablet batch beskrevet. Derudover diskuteres det hvordan, den sikkerhed, der opnås ved et specifikt accept kriterium, kan udledes fra den tilsvarende OC-kurve.

Det vises, hvordan et parametrisk accept kriterium for blandingen eller tabletbatchen, der giver en fastlagt (stor) sikkerhed for at fremtidige stikprøver vil have mindst en fastlagt sandsynlighed for godkendelse under et USP three-class krav.

Egenskaber og robusthed ved ændringsforslag til USPs test for Content Uniformity er belyst ved simuleringer, og enkeltprøvningsplaner for inspektion ved kontinuert variation, der tilstræber at matche USP-forslaget er blevet forslået.

Contents

Preface	iv
Acknowledgements	vi
Summary	ix
Resumé (in Danish)	xi
I	1
1 Background	3
1.1 Outline of the Thesis	5
2 Introduction	7
2.1 Principles of tablet production	7
2.2 Assessment of the uniformity of the blend	10
2.3 Regulatory Affairs	11

2.3.1	Organizations	13
2.3.2	Requirements and Recommendations	15
3	Results and discussion	17
3.1	Variances as a measure of homogeneity	18
3.2	Methods to assess homogeneity and factors that may influence homogeneity	22
3.2.1	The effect of particle size distribution	22
3.2.2	Assessment of homogeneity in specific batches	23
3.2.3	Example	25
3.3	Analysis of acceptance criteria	26
4	Conclusion	31
II	Included papers	35
A	Robustness and power of statistical methods to assess blend homogeneity	37
1	Introduction	39
2	Models of batch homogeneity	40
2.1	The aggregated model	40
2.2	The hierarchical model	41

2.3	Simulated samples	44
3	Assessing factors with influence on the mean content of the active component in a sample	45
3.1	Large and medium scale homogeneity assessed according to the aggregated model (1)	46
3.2	Large and medium scale homogeneity assessed according to the hierarchical model (2)	48
3.3	The effect of including sampling thieves in the model	50
3.4	Conclusion	54
4	Assessing factors with influence on the variation between replicates	55
4.1	Conclusion	59
5	Conclusion	59
 B Comprehensive measures of blend uniformity		63
1	Introduction	65
2	Batches with medium scale variation (variation between areas)	68
2.1	The total variation from the ANOVA table	68
2.2	The variance on a randomly sampled unit from the batch	69
3	Batches with large scale variation (variation between layers)	71
3.1	The total variation from the ANOVA table	73
3.2	The variance on a sample from the batch	75

4	Batches with both large and medium scale variation	75
4.1	The total variation from the ANOVA table	76
4.2	The variance on a sample from the batch	82
4.3	The direct relation between large/medium scale variation and the acceptance criteria	82
5	Discussion and conclusion	82

C On a test for content uniformity in pharmaceutical products

Presented at the First Annual ENBIS Conference, Oslo 2001 **87**

1	Introduction	89
2	The proposed test	90
2.1	Historical notes	90
2.2	Description of the proposed test	91
3	Properties of the proposed test	93
3.1	Description of the OC-surface of the test	94
3.2	“Specification limits” for individual tablets	95
3.3	Details on the effect of individual elements of the test	97
4	Comparison to other test procedures	101
5	Conclusion	102

D	Statistical tests for uniformity of blend/dose samples	105
1	Introduction	107
2	Acceptance criteria and statistical hypothesis testing	111
2.1	Choice of null hypothesis and alternative hypothesis	111
2.2	Confidence intervals and statistical tests	113
3	Notation and distributional assumptions	114
4	Acceptance criteria for the dispersion of doses	116
4.1	Criterion based upon a specified limiting value of sample standard deviation	116
4.2	Criterion based upon a specified limiting value of sample coefficient of variation	118
4.3	Criterion based upon prediction of standard deviation of future samples	123
4.4	A direct approach in terms of population values	129
5	Acceptance criteria with limits on individual measurements	133
5.1	The USP 21 criteria	133
5.2	Three-class attributes and a parametric approach	136
5.3	Confidence region approach	144
5.4	Relation to theories of acceptance sampling by variables	145
5.5	Design of a test with a trapezoidal acceptance region	148
5.6	Discussion	159

6	Assessment of the properties of the USP preview dosage uniformity test	160
6.1	The acceptance value	162
6.2	The simulation study	163
6.3	Robustness against deviation from distributional assumptions	174
6.4	Equivalent single sampling plan	175
7	Further issues	177
8	Discussion	179
9	List of symbols	181

E On particle size distributions and dosage uniformity for low-dose tablets 183

1	Introduction and summary	185
2	Lognormal distribution of particle radii	186
2.1	Distribution of particle radii	187
2.2	Distribution of particle mass for spherical particles	188
3	Distribution of dose content under random mixing of particles	193
3.1	Modelling random mixing	193
3.2	Constant number of particles in tablets	193
3.3	Random variation of number of particles in tablets	195

3.4	Minimum number of particles necessary to secure a specified dose coefficient of variation	200
3.5	Coefficient of skewness and excess for distribution of dose content	202
4	Distribution of dose content under non-random mixing	205
4.1	Spherical particles	209
4.2	Minimum number of particles necessary to secure a specified dose coefficient of variation	210
5	Discussion	210
6	List of symbols	214
 F Case: Analysis of homogeneity in production scale batches		217
1	Purpose	219
2	Experimental	219
3	Statistical Analysis	221
3.1	Assessment of tablet samples	222
3.2	Repeatability	224
3.3	Mean content of the active component	228



Part I

Chapter 1

Background

The background for the Ph.D. project is the general requirements to the pharmaceutical industry of scientific based documentation of the methods used for validation of processes.

To meet the requirements extensive validation testing should be performed at various stages of the manufacturing process to show that various unit operations accomplish what they are supposed to do. The validation testing in the pharmaceutical industry is especially strict compared to requirements in most other industries because failure of meeting a high standard for pharmaceutical products could lead to quite grave consequences.

The pharmaceutical process under consideration in this thesis is the production of tablets which is not an inconsiderable part of pharmaceutical production as an estimated 80% of pharmaceutical products are tablets [1].

Tablets are compacts of powders. Essentially tablets are produced by blending the powdered ingredients until satisfactory uniformity is obtained. Then the tablets are compressed from the powder blend. Hence a critical unit operation in the manufacturing of tablets is the mixing of the final blend. Poor blending or the inability to maintain a blend, i.e. segregation, will inherently lead to problems with the drug content of the tablets compressed from the blend. This is costly in terms of rejected material, extra blending time, and defective end

products. See [2] for a detailed overview over the compliance and science of blend uniformity analysis.

Blend uniformity can be validated by sampling a number of samples from the blend. If the content of the active ingredient in these samples conform to the relevant acceptance criteria the blend is accepted.

Throughout the pharmaceutical industry process validation programs for the manufacturing of tablets have been influenced by the Wolin decision in the U.S. vs. Barr Laboratories [3]. Judge Alfred Wolin defined some of the CGMP (Current Good Manufacturing Practice) requirements for process validation of oral solid dosage forms in greater detail than specified in 21 CFR Part 211 [4]. Particularly it was ruled that the appropriate sample size for content uniformity testing of the final blend in validation and ordinary production batches is up to three times the run weight of the finished product. Larger sample sizes increase the risk of masking insufficient homogeneity on a tablet scale.

The decision caused FDA (The Food and Drug Administration) to reexamine and modify its policies on blend uniformity and sampling techniques. The resulting policies are based on the assumption that current technology provides a means to consistently collect minute representative samples from much larger static powder beds.

However, limitations in the sampling technology makes it difficult to apply scientifically valid methods to blend uniformity validation, because current technology does not provide a method of universally collecting small representative samples from large static powder beds. The problem is the potential for sampling bias. As a result the mean and/or variation between the content in the samples may be significantly different from the mean content / the variation in the blend.

At the moment the tablet press could be viewed as the ultimate sampling device, because the whole batch is being sampled at this stage of the process. However it is not allowed solely to rely on this when demonstrating blend uniformity.

1.1 Outline of the Thesis

This thesis is organized into two parts. The first part is four chapters and contains a description of tablet production and the relevant regulatory affairs as well as a discussion of the results presented in part two.

Part two is six appendices with five articles and manuscripts to articles as well as a case studies. The appendices deals with different aspects of assessment of blend homogeneity and as such they represent the main part of the thesis.

In Appendix A statistical methods to assess blend homogeneity and factors that may have an influence on homogeneity are presented. The focus is on exploratory analysis of blend homogeneity. Appendix F contains a case study using these methods.

In Appendix B two methods to assess the overall (total) variation in the blend are discussed.

The effect of the particle size distribution on the distributed content in blend and tablet samples are discussed in Appendix E.

Finally acceptance criteria for blend and tablet samples are discussed in Appendix C and Appendix D.

The appendices should be read together with the discussion in Chapter 3 as the results are put in a larger perspective in this chapter.

Chapter 2

Introduction

The technical area of this thesis is applied statistics and the field of application is the pharmaceutical industry, specifically tablet production. Thus, the content of the thesis is in the borderland between the areas of applied statistics and pharmaceutical science.

For the reader with a non-pharmaceutical background the principles of tablet production and some relevant concepts are briefly introduced in the following.

2.1 Principles of tablet production

An example of a tablet production is shown schematically in Figure 2.1.

A tablet consists of one or more active ingredients; the drug substance, and some filling materials which have the main purpose to give the tablet suitable physical, biological and chemical properties. For example assure that the drug is released after a certain amount of time in the body or to assure the breaking strength so the tablet does not break into pieces before it is consumed by the patient.

All raw materials are in a powdered form. As the particle size distribution is

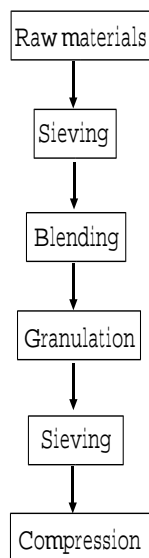


Figure 2.1: Flowchart of tablet production phases.

very important for a number of tablet properties as well as for the variation in content in the tablets the raw materials are initially sieved to eliminate lumps.

Then the raw materials are mixed in a blender until the blend is considered homogeneous. Many different types of blenders exist. The differences may e.g. be due to the physical presentation or due to the mechanical principles used. Further, some blenders may also be used for an eventual granulation of the blend. The purpose of granulation is to obtain particles of more uniform size. This can be done either by breaking larger particles but most often by combining smaller particles to larger particles.

Granulation of the blend is not always necessary, however depending on the properties of the blend granulation may result in improved floating properties which is important at the tablet press. Further, by granulation the state of blend is partly fixed thus reducing the risk of deblending, etc.

The granulate may be sieved again to eliminate lumps. Then the blend is compressed at the tablet press. Some products are produced by direct compression of the blend without granulation.

The process in the tablet press is in principle a successive series of dosings of the blend or granule to the die in the tablet press, whereupon each dose is compressed to a tablet by compaction between two punches.

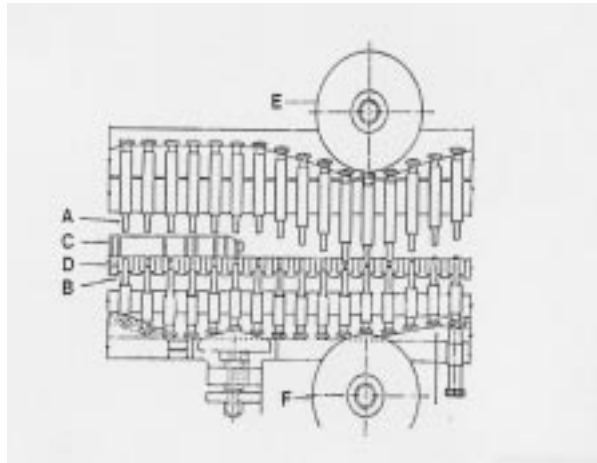


Figure 2.2: Tablet press

Figure 2.2 shows a row of punch stations on the tablet press. A and B are upper and lower stamp respectively. C is the place where the blend is led to the press. Beneath this place the matrix (D) passes at the same time as the dosing takes place. Then the filled matrices and the corresponding punches passes two rolls (E and F) that presses the punches together. Immediately after this the tablet is ejected and pushed away.

After the compression at the tablet press the tablets may be coated for example to protect the active ingredient from decomposition due to light or moisture or to facilitate packaging, mask unpleasant taste or smell etc.

2.2 Assessment of the uniformity of the blend

The content uniformity of the tablets is an important quality measure of the final product. As the content uniformity is closely related to the uniformity of the blend, it is important to monitor blend uniformity. Inhomogeneities of a blend can either be due to insufficient mixing or to deblending under transportation or storage of the blend.

In practice blend uniformity is assessed by collecting a number of samples from the blend, each sample being of the size of 1-3 times the corresponding tablets. The sampling locations must be carefully chosen to provide a representative cross-section of the granulation. The resulting samples are then assayed using the same methods used to analyze the finished product. Content uniformity is established if the drug content of the samples conform to a predetermined criterion.

The current state of the art regarding sampling technology is a device referred to as a sampling thief. Many different types of sampling thieves have been developed. However, in general a sampling thief consists of two concentric tubes. The inner tube is solid except for one or more chambers that allow for sample collection. The outer tube is hollow and contains openings that can align with the chambers in the inner tube. A handle, located at the top of the thief is used to rotate the inner tube within the outer tube in order to open or close the thief. A sample is collected by inserting the closed thief into the blend. Then the handle is rotated in order to open the thief allowing the sample to flow into the sampling chamber in the inner tube. Then the thief is closed and pulled out of the blend.

Figure 2.3 shows examples of two different types of sampling thieves. In these thieves the sampling chamber is located at the tip of the thief rather than on the side as described above.

However, the two “Golden Rules” of sampling [5] states that a powder should be sampled when in motion and that the whole of the stream of powder should be taken for many short increments of time. Any sampling methods which does not adhere to these rules should be regarded as a second-best method liable to lead to errors.

Collecting blend samples by the use of sampling thieves violates these Golden

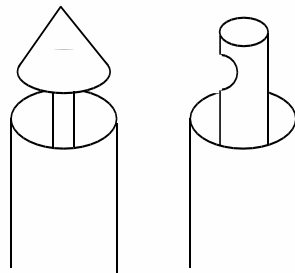


Figure 2.3: Two examples of sampling thieves. In these thieves the sampling chamber is located in the tip of the thief.

Rules, and the result is the risk of sampling error or bias. The presence and the size of such sampling errors and bias depends on factors such as sampling device, sampling technique, blend formulation, blender size, sample location and size of the collected sample. For a more detailed discussion of these factors see e.g. [2].

Figure 2.4 shows a boxplot for the results of samples from blend and the corresponding tablets for seven batches produced at Pilot Plant, Novo Nordisk A/S. Each batch corresponds to a different value of label claim (LC). It is seen that for all batches the mean content in blend samples is larger than the mean content in the tablet samples. This could be a typical result of sampling bias.

2.3 Regulatory Affairs

Pharmaceutical companies often sell their products in several parts of the world. In order to do that the companies have to comply with all the requirements covering the countries or areas in which they sell their products. In order to reduce the need to duplicate the testing carried out during the research and development of new medicines efforts are done at making greater harmonisation in the interpretation and application of technical guidelines and requirements for product registration. This work is particularly organized via The International Conference on Harmonisation, ICH, which is a joint initiative between

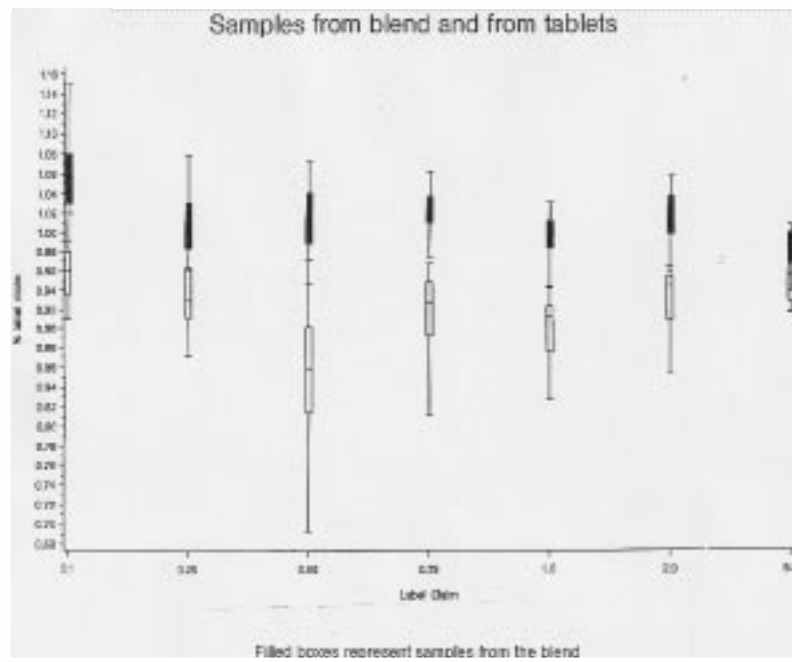


Figure 2.4: Samples from blend and tablets.

the regulatory authorities of Europe, Japan and the United States and experts of the pharmaceutical industry in the three regions. Until this harmonization is completed the pharmaceutical companies have to comply with various sets of requirements and in this situation the American legislation is very important because for many pharmaceutical companies the American market is one of the most important markets. Therefore the focus in this thesis is on rules and requirements in the American legislation.

2.3.1 Organizations

The main actors in the U.S. are the FDA, USP, PDA and PQRI. In the following these organizations are introduced.

The Food and Drug Administration, FDA

Most countries have a governmental drug administration which approves drug products. In the U.S. the drug administrative organ is the FDA. FDA is an agency, charged with protecting American consumers by enforcing the Federal Food, Drug, and Cosmetic Act and several related public health laws. Among other things it monitors the manufacture, import, transport, storage and sale of medicines and medical devices.

In deciding whether to approve new drugs, FDA does not itself do research, but rather examines the results of studies done by the manufacturer. A part of this investigation is to assess whether the new drug produces the benefits it is supposed to without causing side effects that would outweigh those benefits [6].

The United States Pharmacopeia, USP

USP is the American pharmacopoeia responsible for developing public standards and information concerning public health.

In pursuit of its mission to promote public health, USP establishes standards to

ensure the quality of medicines for human and veterinary use. Manufacturers must meet these standards to sell their products in the U.S. The standards are officially recognized standards of quality and authoritative information for the manufacturing and use of medicines and other health care technologies [7].

Parental Drug Association, PDA

PDA is a non-profit international association involved in the development, manufacture, quality control and regulation of pharmaceuticals and related products. PDA is a leading technical organization in the fields of parental science and technology that tries to influence the future course of pharmaceutical products technology.

The mission is to support the advancement of pharmaceutical technology by promoting scientifically sound and practical technical information and education for industry and regulatory agencies [8].

The Product Quality Research Institute, PQRI

PQRI is designed to provide a neutral environment where FDA, academia and industry can collaborate on pharmaceutical product quality research and develop information in support of policy relating to regulation of drug products. PQRI supports the priorities of FDA to improve and enhance its science base and provides scientific evidence for policy enactment or changes. PQRI also serves the pharmaceutical industry by promoting efficiency and consistency in the regulatory processes.

A number of working groups are established. The ultimate goal of these working groups is to develop scientific knowledge that will result in appropriate changes to regulatory policies to make them less burdensome [9].

2.3.2 Requirements and Recommendations

A very important document is 21 Code of Federal Regulations. 21 CFR is a very general law describing current good manufacturing practice (CGMP)¹. Of special interest is 21 CFR Part 210 and Part 211 [4] describing respectively processing, packing, or holding of drugs (part 210) and for finished pharmaceuticals (part 211).

21 CFR is published by FDA. Pharmaceutical companies on the American market have to comply with this law. As the law is very general it does not give many specific technical details on how to comply with the law. Some of these details are found in the current American pharmacopoeia, USP 24. As an example the pharmacopoeia specify how to perform content uniformity testing, i.e. how to test the uniformity of tablets. Also a lot of guidance documents and guidelines on various topics are published by FDA. The content of these documents are not directly 'law' but they contain detailed information on how FDA interpret the law and more detailed suggestions and recommendations on what the manufactures can do if they want a drug to be approved. As an example [11] gives guidelines on blend uniformity testing. For a more detailed description of these documents see e.g. [2].

In 1996 FDA proposed some changes to 21 CFR Part 210 and Part 211. Regarding blend uniformity testing the most important change is a new paragraph 211.110(d) that specifically require blend samples to approximate the dosage size of the product for blend uniformity analysis. Thus, this proposed amendment would for the first time legally oblige the pharmaceutical industry to conduct blend uniformity analysis using unit dose testing.

¹CGMP regulations are based on fundamental concepts of quality assurance: (1) Quality, safety, and effectiveness must be designed and built into a product; (2) quality cannot be inspected or tested into a finished product; and (3) each step of the manufacturing process must be controlled to maximize the likelihood that the finished product will be acceptable. [10]

Chapter 3

Results and discussion

With a background in the legal requirements for the pharmaceutical industry to validate critical unit operations as for example the mixing of the final blend in the tablet production, this thesis addresses some of the problems related to assessing the homogeneity in powder blends.

Before starting the production of a new product or changing an existing blending or blend sampling process it is important to investigate factors that may have an influence on the processes. For this kind of exploratory investigations it is meaningful not just to evaluate the overall homogeneity but to consider homogeneity on different scales in the blend. More specific in this thesis the homogeneity is evaluated on a large, a medium and a small scale. Such an evaluation on more than one scale will enhance the understanding of the processes. Statistical methods to assess blend homogeneity on different scales and to evaluate factors that have a possible influence on the homogeneity are presented in Appendix A. An example of an explorative analysis is given in Appendix F. Even though the number of actually conducted experiments in this example was smaller than originally planned and therefore the resulting design is not 'optimal' for the statistical methods used this experiment has been chosen as an example, as it includes both blend and tablet samples. Comparing blend and tablet samples is a more holistic approach than analysing blend and tablet results separately. The example should be seen as a guidance on considerations and conclusions with relevance for this type of analysis.

Regarding the patients using the final tablets it is of less relevance if the variation between the doses is due to large, medium or small scale variation in the blend. In this relation the magnitude of the total variation in the batch of tablets is relevant. The total variation between the content of the tablets is closely related to the total variation in the blend. Therefore for practical purposes it is relevant to control the total variation in the blend. In Appendix B the three scales of homogeneity discussed in Appendix A are related to overall measures of blend homogeneity. The measures of overall homogeneity are compared by relating them to an acceptance criterion for blend uniformity.

Acceptance criteria for both blend and tablets are usually assessed assuming a normal distribution of content in the samples. However, actual distributions of particle sizes are often seen to be skewed. This might have an effect on the shape of the distribution of content in blend and tablet samples. Therefore, in Appendix E the effect of a skewed particle size distribution on the distribution of content in the samples is discussed.

Keeping in mind, that for example a skewed particle size distribution can influence the distribution of the content in the blend and tablet samples, the statistical properties of acceptance criteria for blend and tablet samples are discussed under the normal assumption in Appendix C and Appendix D. Appendix C gives background and preliminary considerations to the analysis in Appendix D. Further, the acceptance criteria analysed in Appendix C is an earlier version of the corresponding acceptance criteria analysed in Appendix D.

In the following the results and discussions of these are given in more detail.

3.1 Variances as a measure of homogeneity

It comes natural to think of homogeneity as some kind of variance being small. However, even though variation is an often used parameter in various relations it is not straight forward in case of bulk materials to define homogeneity as a variance. The reason is that bulk materials essentially are continuous and do not consist of discrete, identifiable, unique units or items, i.e. there is no natural unit or amount of material that may be drawn into the sample [12].

A single particle is not a suitable unit as it is too small for practical purposes.

Rather, the ultimate sampling unit must be created, at the time of sampling, by means of some sampling device. The size and form of the units depend on the particular device employed, how it is used, the nature, condition, and structure of the material, and other factors.

However, this definition of a unit is convenient and conceptual and further for practical purposes the size of a sample do not differ much from the size of a tablet produced from the blend. Thus, a unit defined in this way is in agreement with the tableting process and therefore makes it less complicated to compare homogeneity in the blend to homogeneity in the tablets.

By adapting a sample as a definition of a unit the variance between the drug content in a number of units can be calculated and used as a measure of homogeneity.

When a unit has been defined the next problem is to decide where to sample and how many samples to collect to be able to estimate a variance that is representative for the blend homogeneity. In this relation it should be mentioned that as an example the total amount of drug in a 360 kg batch (drug and filling material) could be as little as 0.5 kg, and the weight of a sample less than f.ex. 80 mg. With these orders of magnitude and in case of batch inhomogeneity a variance estimated between samples sampled close to each other differs from a variance estimated from samples collected far apart. Hence, for exploratory purposes it is relevant to assess different types of variances, i.e. variances estimated from samples sampled closely and variances based on samples sampled far apart.

In Appendix A a model that describes blend inhomogeneity (variation between sample 'units') on three scales is introduced. The three scales are referred to as small, medium and large scale variation and they correspond to respectively variation between the content in neighbouring samples/replicates, variation between the mean content in areas within a layer in the blend and variation between the mean content in different layers in the blend. In statistical terms this is a hierarchical or a nested model. In Appendix A large scale variation refers to inhomogeneities between layers in the blend as vertical inhomogeneity is a very likely result of deblending. However, in case of suspicion of inhomogeneity in the horizontal direction the model could easily be changed to model this kind of inhomogeneity. Further, the hierarchical model can also be changed into modelling inhomogeneity on e.g. four or two scales of inhomogeneity if

this seems to be more relevant.

In case of blend homogeneity the large and the medium scale variation (measuring differences between the mean content in respectively layers and areas within a layer) are zero. The small scale variation is an inherent variation in the blend and therefore it is not zero in case of homogeneity. However, in case of homogeneity the small scale variation is independent of in which layer of the batch it is estimated.

It should be noted that in the literature several examples exist of models taking into account correlation between the samples measured as a function of the distance between the spots in the blend from which the samples are collected. (See e.g. [13]). However, these models are generally not used in practice. With future techniques as e.g. NIR (near infra red) techniques correlation as a function of distance may be used in relation to image analysis methods. However, NIR technology is not commonly introduced in production yet, and the focus of this thesis is to develop and improve methods to assess uniformity within the scope of current sampling technology, the sampling thieves.

For explorative purposes assessing inhomogeneity on different scales is relevant. However, when it comes to the patients using the tablets a single measure of the overall homogeneity in the blend is relevant as the overall blend homogeneity corresponds to the overall homogeneity of the content in the tablets. In Appendix B two methods of measuring the overall variation in the blend is discussed under the assumption that homogeneity can be modelled by the hierarchical model presented in Appendix A. Both methods relate the overall variation to the variation measured on the three scales of homogeneity defined in Appendix A.

The first method is to use the total variation from the analysis of variance (ANOVA) table corresponding to the hierarchical model for the variation in the batch as an estimate for the overall variation. The other method is to use the total variation on a randomly collected sample from the blend as an estimate of the overall variation in the blend.

The difference between the estimates of the overall variation obtained with each of these two methods depends on the sampling plan used to collect the samples on which the estimates are based.

If a patient only uses one randomly sampled tablet for example when taking an aspirin to relieve the pain of a headache, he/she will experience a deviation from LC corresponding to the variance on a randomly chosen tablet. However, if the patient uses more than one tablet as part of an ongoing treatment, the total variation in drug content experienced may depend on the way the tablets are collected. Are they randomly chosen from the batch or do they all come from the same part of the batch etc. The tablets in a single package will in general not be sampled from a balanced, hierarchical sampling plan as in Appendix A and Appendix B, and even if the tablets by accident were sampled in accordance with a hierarchical sampling plan, the "sampling plan" would be unknown. Hence, regarding the total variation experienced by a patient using more than one tablet neither method of estimating the overall variation is ideal.

Another criteria for deciding which estimate to use as a measure for the overall variation in the blend is to consider the properties of the acceptance criteria for the blend. Acceptance criteria are discussed in more detail in Section 3.3. In case of uncorrelated samples, which corresponds to the model in Appendix A with no variation between layers and a hierarchical sampling plan with only one replicate per area, the two measures for the overall variation in the blend are identical. Otherwise the measure of the total variation corresponding to the ANOVA table in general leads to more efficient and less ambiguous properties of the acceptance criteria for blend homogeneity.

In conclusion variance can be used as a measure of (in)homogeneity. For explorative purposes it is relevant to look at variances at different scales. In other situations an overall measure of the batch homogeneity may be more convenient and relevant. Two methods to estimate the overall variance are presented. None of these truly describes the total variation experienced by a patient using more than one tablet - but it is very complicated if possible at all, to estimate this total variation. However, regarding acceptance criteria for blend uniformity the total variation from the ANOVA table is relevant.

3.2 Methods to assess homogeneity and factors that may influence homogeneity

In the thesis two different approaches to assess homogeneity and factors that may influence homogeneity have been introduced.

The first approach described in Appendix E leads to a model of the best obtainable blend and content uniformity derived from the distribution of particle radii. However, the best obtainable homogeneity is a 'theoretical' limit that holds for all batches with the same distribution of particle radii, and therefore this approach does not lead to information on the actual homogeneity of a given batch. The second approach described in Appendix A introduces two methods to assess the homogeneity of a specific batch.

3.2.1 The effect of particle size distribution

Particle size distributions are often seen to be skewed and it has been shown in Appendix E that this feature affects the distribution of content in blend and tablet samples.

For a log-normal distribution of particle diameters, the resulting distribution of particle mass (volume) is also a log-normal distribution. It is found that skewness and excess (heavy-tailedness) of the distribution of particle radii is amplified when transformed to particle mass. The larger the coefficient of variation in the distribution of particle radii the more pronounced the amplification. The relation between the coefficient of variation for particle mass and the coefficient of variation for particle radii is given in a table.

Beside the variation in particle mass the variation in dose content is affected by variation in the number of particle in a sample. For a homogeneous blend with a random scattering of particles over the blend it is demonstrated that for a given distribution of particle sizes the variation in the distribution of absolute doses is proportional to the average number of particles in the samples (tablets). Further, it is shown that the larger the average number of particles in the sample the closer the distribution of content in the samples is to a normal-distribution.

For spherical particles an explicit relation between the variation in the relative doses and the mean and the coefficient of variation in the distribution of particle radii is given.

3.2.2 Assessment of homogeneity in specific batches

Two methods are introduced in Appendix A to assess homogeneity and factors that may influence homogeneity in a specific blend. The two methods are based on respectively Generalized Linear Models (GENMOD) and General Linear Models (GLM). General Linear Models can be used to assess differences in mean content in respectively layers and areas within a layer. Generalized Linear Models are here used to assess differences in variance. More specific to assess whether the size of the small scale variation is constant throughout the batch.

In practical applications a Generalized Linear Model should be applied first to assess if the small scale variation/variation between replicates is constant throughout the blend. If this is not so, it should be accounted for in the General Linear Model.

GENMOD

In Appendix A a Generalized Linear Model is used to assess the influence of layers on the small scale variation. For samples simulated from a hierarchical model with three layers, four areas within each layer, and three replicates within each area the following was found: For the 5% level test the standard deviation between replicates within an area has to be 4.5 times larger in one layer than the standard deviation corresponding to replicates within an area in another layer for the effect of layers to be declared significant with a probability of at least 0.95.

However, depending on the experimental design and the assumptions made, the method presented in this section can also be used to assess factors influencing sample error (variation). For example the method can be used to test if one sampling thief leads to larger variation between replicate samples than another thief.

Using the thief leading to the smallest variation between replicates will reduce the risk of incorrectly rejecting a batch because of suspicion of inhomogeneity. Other external sources of variation (as e.g. the sampling procedure) may have a similar effect on the small scale variation. Examples of such analysis are given in Appendix F.

GLM

When a Generalized Linear Model has been applied a General Linear Model can be applied specially to assess the medium and the large scale variation as well as factors that may influence these types of variation. In case the variation between replicates is found not to be constant throughout the blend, this should be corrected for in the analysis by introducing appropriate weights

Under the assumption that the variation between replicate samples is independent of the layer and that there is no interaction between the factors in the model, two statistical methods to describe blend homogeneity have been investigated.

The first statistical method (using the aggregated model) corresponds to an 'aggregated' definition of homogeneity in the sense that large and medium scale variation in the batch is assessed as a whole.

The other statistical method (using the hierarchical model) corresponds to a homogeneity definition with two different criteria; one explicitly regarding the large scale variation and the other explicitly regarding the medium scale variation.

The analysis showed that the two methods are approximately equally good at detecting inhomogeneity. That is, an analysis according to the aggregated model can be used to detect inhomogeneity even in situations with large scale variation (variation between layers).

The most important difference between the two types of analysis is that when inhomogeneity is declared according to the aggregated analysis the result does not reveal whether this inhomogeneity is due to large or medium scale variation in the batch. However, the hierarchical model explicitly assesses respectively large and medium scale variation.

The power of the respective tests as a function of the standard deviation corresponding to respectively variation between layers and variation between areas within a layer are shown in Figure 3 to Figure 5 in Appendix A. The standard deviation corresponding to respectively variation between layers and variation between areas within a layer are measured relatively to the standard deviation corresponding to replicates.

Finally, for the given design (and for $\sigma_{rep} = 1$) the power of the test of a thief effect was shown in Figure 6 in Appendix A. The test of the thief effect is independent of σ_{layer} and $\sigma_{area,hi}$.

With the given design (and for $\sigma_{rep} = 1$) a difference greater than 1.5 in mean content in samples from two different thieves will be detected with a probability of at least 0.95.

3.2.3 Example

As an example of how the robustness and power of a test can be used to evaluate the test result the test of the small scale variation in Appendix F is discussed in relation to the results from Appendix A.

As the resulting design in Appendix F is neither balanced nor identical to the design from which the robustness and power are assessed, the example should be seen as a guidance on the type of considerations to make when evaluating test results.

Variation between replicate samples

In Appendix F the variation between replicates, σ_{rep}^2 , tends to be larger the lower the layer the samples are sampled from. However, the tendency is not significant.

The estimated variance components, σ_{rep}^2 , are given in the Table 3.1. The estimates in the table are multiplied with 1000 compared to the results in Case F. The reason is that in Appendix A the unit is % LC rather than fraction LC.

	Batch 1			Batch 1		
Thief	1	2	3	1	2	3
$\sigma_{rep,top}^2$	0	0	1	0	3	8
$\sigma_{rep,bottom}^2$	1	7	17	4	35	98

Table 3.1: Estimates of variance components from Appendix F.

It is seen from the three cases in the table where $\sigma_{rep,top}^2 \neq 0$ that the estimate of $\sigma_{rep,bottom}^2$ is between 10 and 20 times as large as $\sigma_{rep,top}^2$. At first a difference that large may be expected to be found significant. However, for the design in Appendix A it was showed that σ_{rep} in one layer should be more than 4.5 times as large as σ_{rep} in an other layer for the test to show significance with a high probability. This corresponds to one variance estimate being $4.5^2 \approx 20$ times larger than the smallest. This result is valid for the balanced design with 12 pairs of replicates in each batch and analysed with a model with one factor (layer).

The design in Case F is not balanced, it has only 6 pairs of replicates in each batch and two factors included in the model (layer and thief). Hence, with an estimate of $\sigma_{rep,bottom}^2$ being 10 to 20 times as large as $\sigma_{rep,top}^2$ it is not surprising that the test shows no significance - however, the estimated difference between $\sigma_{rep,top}^2$ and $\sigma_{rep,bottom}^2$ may still be real and give practical and valuable insight to the sampling process.

3.3 Analysis of acceptance criteria

When a parameter expressing the overall (total) (in)homogeneity of the blend has been estimated the batch quality can be evaluated by comparing the parameter estimate to some acceptance criteria or critical value. The critical value could be determined from some theoretical model of the ultimate limit of homogeneity. These theoretic limits could for example be the variance in a completely ordered or in a completely random blend. The most common definition of a perfectly random blend is one in which the probability of finding a particle of a constituent of the blend is the same for all points in the blend. More than 30 different criteria relating the sample variance to theoretical limits have been proposed by various investigators [14]. These criteria are referred to as mixing indices in the literature. The analysis of variance method e.g. presented in Ap-

pendix A could also be taken as basis for a mixing index (see also [15]). In this case the theoretical limit of homogeneity is the various variance components being zero. It should be noted that a variance component being zero serves more as a theoretical value for homogeneity. It is not useful as an acceptance criteria. Further, it should be noted that in a homogeneous blend the large and medium scale variation are negligible. However, the small scale variation is an inherent variation that is non-zero even in a random blend.

Alternatively to the models for ultimate limits of homogeneity the quality of the blend can be evaluated in accordance to some practical criteria assessing if the homogeneity is satisfactory for the blend to serve its purpose.

The properties of such acceptance criteria can be investigated as a function of e.g. the true mean and total variation in the batch similarly to the analysis of the properties of the acceptance criteria in Appendix C and Appendix D.

The discussion of the acceptance criteria and the derivation of expressions for acceptance probabilities in the appendices is performed under the assumption that individual sample values may be represented by independent, identically distributed variables and that the distribution of sample results may be described by a normal distribution.

Thus, when the samples are tablets selected from a batch, the assumption corresponds to assuming that the overall distribution of dose content in the batch may be represented by a normal distribution and that individual dosage units are selected at random from the dosage units in the batch. When the samples are blend samples, the assumption analogously corresponds to assuming that the overall distribution of such potential samples from the blend may be represented by a normal distribution and that samples are taken at randomly selected positions in the blend. Thus, the model will not be adequate when the overall distribution in the blend (or batch) is bimodal or multimodal corresponding e.g. to stratification, when the distribution is skewed, e.g. as a result of deblending, or when the distribution has heavier tails than the normal distribution, e.g. as a result of imperfect mixing (clustering) or of using drug particles that are too large for the intended dosage (see Appendix E).

When sampling is performed under a hierarchical (or nested) scheme as suggested e.g. by PQRI, the model will only be adequate in such (rather unlikely) situations where there is no correlation between subsamples from the same

location in the blend (see [16]).

However, even despite these restrictions the mathematical analytical discussion serves a purpose of clarifying and illustrating the statistical issues involved, thereby providing further insight in the properties of various tests that have been proposed in the pharmaceutical literature.

Under the assumption that the distribution of sample results may be described by a normal distribution, the following results regarding acceptance criteria have been derived in Appendix C and Appendix D.

In essence the purpose of using acceptance criteria is to secure a certain quality of the product under concern. Thus in industrial or commercial practice, product requirements are often formulated as specifications for individual units of product, but may also include specifications for such batch or process characteristics as batch fraction nonconforming or standard deviation between units in the batch.

However, regulatory practice for pharmaceutical products has most often specified criteria for sample values rather than providing specifications for the entity under test. As therefore control and monitoring procedures in tablet production are based upon samples from the blend, or from the batch of tablets, there is an inherent uncertainty concerning the actual dispersion in the blend or batch being sampled. This uncertainty is partly due to sampling and partly due to the (in)homogeneity of the blend/batch.

The statistical tool used to link sample result and acceptance criteria to the actual dispersion in the blend or batch is an OC-curve (or surface) that shows the probability of passing the acceptance criteria as a function of e.g. fraction nonconforming or true mean and standard deviation in the blend or batch, as an OC-curve (or surface) reflects the effect of such sampling uncertainty.

When properties of an acceptance criteria have been described through the corresponding OC-curve the next issue is to determine the assurance related to the acceptance criteria. This assurance can also be determined from the OC-curve.

In Appendix D statistical tools and methods that can be used to determine how assurance depends on sample size (i.e. how to set up a criterion that gives a

certain assurance with an appropriate sample size) are described and discussed for simple acceptance criteria (sample standard deviation and coefficient of variation as well as an USP criterion that includes a test by attributes).

Also in the appendix it is shown that a three-class attribute criteria as e.g. in USP 24 for content uniformity in essence controls the proportion of tablet samples outside the inner set of limits for individual observations. For normally distributed observations this is identical to control the combination of batch mean and standard deviation, i.e. a parametric acceptance criterion. It is shown how to set up parametric acceptance criteria for the batch that gives a high confidence that future samples with a probability larger than a specified value passes the USP three-class criteria.

In the literature changes to the procedure in USP have been proposed. In general the proposed test procedure is similar to the parametric criteria mentioned above. In the thesis simulations have been performed both for normally and log-normally distributed content in the tablets. The simulations revealed that the test is relatively robust to deviations from the normal distribution. This is relevant as such deviations for example is seen in case of low-dose tablets with large particle radii as also discussed in the thesis.

Finally single sampling acceptance plans for inspection by variables that aim at matching the USP proposal have been suggested.

Chapter 4

Conclusion

In this thesis the use of statistical methods to address some of the problems related to assessment of the homogeneity of powder blends in tablet production is discussed.

When assessing homogeneity the first problem is how to define homogeneity of the blend. This is not straight forward as bulk materials have no natural unit or amount of material that may be drawn into a sample. However, a blend sample of the size of one to three times the size of a tablet is a convenient unit.

With this definition of a unit, variances between blend samples can be used as a measure of the (in)homogeneity in a blend.

In the thesis a hierarchical (or nested) as well as an aggregated model has been introduced to describe (in)homogeneity. The hierarchical model specifically takes into account deblending in a specified direction. Both the hierarchical and the aggregated model can be used to detect inhomogeneity. However, in case of inhomogeneity the hierarchical model provides the most detailed information on the type of inhomogeneity.

Regarding the end users of the tablets the total variation between the tablet content is relevant. This variation is closely related to the overall variation in the blend. Two methods to determine the overall variation in the blend have

been suggested. One of the methods (estimating the overall variation from the ANOVA table) leads to less ambiguous properties of acceptance criteria for blend uniformity. However, none of the methods truly describes the variation experienced by a patient, as this variation depends on how the tablets in the package are selected from the batch of tablets.

It has been shown that particle size distribution may have an influence on the distribution of content in blend and tablet samples. Specially for low-dose tablets it is important to keep the particle radii small (and the number of particles large) to minimize the variation in content in the blend and tablet samples.

Assuming perfect mixing, and a log-normal distribution of particle sizes, the requirement on the coefficient of variation in the distribution of dosage units is essentially a general requirement on the minimum average number of particles in a dosage unit. This minimum average number of particles does not depend on label claim. However, as the average number of particles in tablets depend on label claim, a blend that might produce a satisfactory distribution of doses (in terms of the coefficient of variation in the distribution of relative doses) for large dose tablets need not be satisfactory for smaller dose tablets.

Interpreting the results in terms of blend samples rather than samples of tablets from a batch, it is of interest to note that the practical necessity of using blend samples that are larger than the dosage units imply that the coefficient of variation in such blend samples is smaller than the coefficient of variation in the resulting dosage units. For blend samples that are four times the size of the final dosage units, the coefficient of variation in the blend samples is only half the size of the coefficient of variation in the final dosage units. Moreover, a larger blend sample might mask departure from normality in the distribution of dose content in low-dose tablets.

Generalized Linear Models (GENMOD) can be used to assess factors that may have an influence on a variance (e.g. the effect of layers on the replicate variance). General Linear Models (GLM) can be used to assess factors that may have an influence on the mean content in blend samples, e.g. a sampling device leading to sampling bias. For a specific sampling design the power and robustness of the statistical tests related to the GENMOD and GLM models have been assessed.

A central problem is to develop acceptance criteria for blends and tablet batches

to decide whether the blend or batch is sufficiently homogeneous to meet the need of the end users. Under the assumption that the content in blend and tablet samples is normally distributed properties of a number of acceptance criteria have been discussed.

Regulatory practice related to tablet production are most often criteria specifying limits for sample values rather than for the actual homogeneity in the blend or batch of tablets. This leads to an inherent uncertainty concerning the homogeneity in the blend or tablet batch. This uncertainty is partly due to sampling and partly due to (in)homogeneity of the blend or batch.

In the thesis it is shown how to link sampling result and acceptance criteria to the actual quality (homogeneity) of the blend or tablet batch. Further it is discussed how the assurance related to a specific acceptance criteria can be obtained from the corresponding OC-curve.

Also in the thesis it is shown that a three-class attribute criteria as e.g. in USP 24 for content uniformity in essence controls the proportion of tablet samples outside the inner set of limits for individual observations. For normally distributed observations this is identical to control the combination of batch mean and standard deviation, i.e. a parametric acceptance criterion. It is shown how to set up parametric acceptance criteria for the batch that gives a high confidence that future samples with a probability larger than a specified value passes the USP three-class criteria.

In the literature changes to the procedure in USP have been proposed. In general the proposed test procedure is similar to the parametric criteria mentioned above. In the thesis simulations have been performed both for normally and log-normally distributed content in the tablets. The simulations revealed that the test is relatively robust to deviations from the normal distribution. This is relevant as such deviations for example is seen in case of low-dose tablets with large particle radii as also discussed in the thesis.



Part II

Included papers

A

Paper A

Robustness and power of statistical methods to assess blend homogeneity

1 Introduction

In this appendix the robustness and power of statistical methods to assess blend homogeneity is discussed. The statistical methods assess various factors that may alter batch homogeneity and therefore one could also say that the statistical methods assess blend inhomogeneity. For convenience the term homogeneity as well as the term inhomogeneity will be used in relation to the statistical models. The discussion is based on simulations of blend samples in SAS [17]. However, before simulations of the blend samples are made it is necessary to have a definition and a model for blend homogeneity.

Two different models for blend homogeneity are presented: An overall or aggregated model that does not distinguish between horizontal and vertical (in)homogeneity and a hierarchical model that specifically takes into account the situation where deblending or insufficient mixing causes inhomogeneity in the vertical direction. This model also accounts for (in)homogeneity in the horizontal direction. The aggregated and the hierarchical models assess large scale and medium scale variation¹. Small scale variation is assessed in a separate analysis.

All samples are simulated in accordance with the hierarchical model. However, both the aggregated and the hierarchical model are used in the statistical analysis of the simulated samples.

Finally, a number of batches are simulated in accordance with a hierarchical model including small scale variation, and a statistical method specifically useful to assess small scale variation is presented.

¹The terms large, medium and small scale variation should be taken loosely. They are meant as a convenient way to distinguish between different types of (in)homogeneities in a batch. By large scale variation is meant variation between samples collected so far apart that the distance between them should be measured on a 'large scale', e.g. top, middle and bottom layer of the batch. Similarly small scale variation means variation between samples collected so close to each other (e.g. replicate samples) that the distance between them should be measured on a 'small scale'. However, it is very important to understand that large and small scale does not refer to the size of the the respective variations. Thus, the large scale variation could actually be smaller than the variation measured on the small scale.

2 Models of batch homogeneity

In the following two different statistical models of batch homogeneity are introduced: the aggregated and the hierarchical models.

2.1 The aggregated model

No definitive definition of homogeneity of a powder blend exists. One definition relates to the acceptance criteria for blend uniformity analysis as suggested by the FDA in a Draft Guidance [11] as well as the the first stage of the blend sample criteria suggested by PQRI [18]. In essence these acceptance criteria controls the variation between a number of locations (areas) in the batch. PQRI [18] explicitly recommends to assess this variation on the basis of samples from at least ten different areas with at least three replicates from each area.

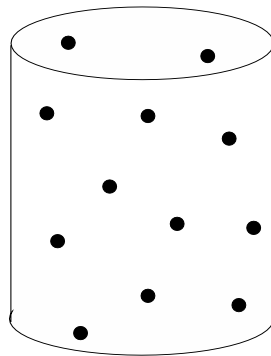


Fig. 1: A batch in which twelve areas are identified for sampling.

Figure 1 shows an example of a batch with twelve areas representatively (and randomly) distributed in the batch. Sampling three replicates from each of the twelve areas would be in compliance with the recommendations from PQRI.

The variation between samples collected in accordance with this sampling plan

can be modelled by the following statistical model:

$$Y_{ij} = \mu + A_i + E_{ij}, \quad \begin{cases} i = 1, \dots, 12 \\ j = 1, 2, 3 \end{cases} . \quad (1)$$

Y_{ij} is the content of a sample, μ the mean content in the batch (including bias due to sampling and chemical analysis), A_i the effect corresponding to the area from which the sample is taken, and E_{ij} is the error term corresponding to the j 'th replicate from the i 'th area. Further $A_i \in N(0, \sigma_{area,agg}^2)$ and $E_{ij} \in N(0, \sigma_{rep}^2)$. The variation, $\sigma_{area,agg}^2$, can be thought of as an aggregate of the large and the medium scale variation in the batch.

If all samples are collected in random order the design is a simple single-factor design. The experimental design, the analysis, assumptions and the statistical terminology related to such a design is given by Montgomery [19].

2.2 The hierarchical model

When basing the assessment of batch homogeneity on a set of blend samples it is important that the samples are representative, i.e. randomly selected locations. However, according to e.g. FDA [20] it is also important that the specific areas of the blender which have the greatest potential to be non-uniform is represented. One common type of inhomogeneity is that the mean content in the top, middle and bottom layer of a batch differ due to deblending or insufficient mixing. Hence, in these cases it is important to collect samples from all layers. Several examples of sampling plans that divide the batch into layers are seen in the literature (see e.g. Kræmer [21] and Berman [5]). Further FDA [20] states that for drum sampling samples should be collected from three layers of the drum: the top, middle and bottom layer.

Figure 2 shows an example of a batch divided into three layers. Within each layer samples are taken from the points of a triangle, as well as from the centre, thus four areas within each layer are chosen. By changing to the next level, the triangle is rotated by 180° . In the left corner of the figure a top view of the sampling scheme is shown.

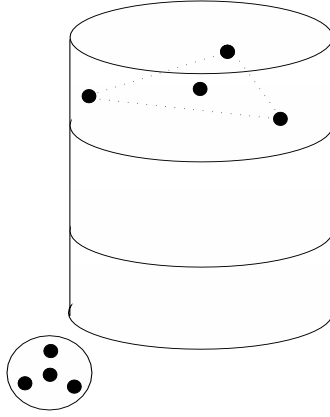


Fig. 2: A batch divided in three layers: top, middle and bottom. Variation between the mean content in the layers represents large scale variation. Variation between the mean content in the four areas within a layer represents medium scale variation. To the left a top view of the sampling scheme is seen.

Variation between the mean content in the three layers can also be thought of as large scale variation in the vertical direction. The aggregated model (1) does not explicitly account for this type of variation or inhomogeneity. However, it is accounted for in the following model:

$$Y_{ijk} = \mu + L_i + A(L)_{j(i)} + E_{ijk}, \quad \begin{cases} i = 1, 2, 3 \\ j = 1, 2, 3, 4 \\ k = 1, 2, 3 \end{cases} \quad (2)$$

L_i is the effect from layer i , $A(L)_{j(i)}$ is the effect from the j 'th area within layer i and E_{ijk} is the effect from the k 'th replicate from area j in layer i . Further it is assumed that $L_i \in N(0, \sigma_{layer}^2)$, $A(L)_{j(i)} \in N(0, \sigma_{area,hi}^2)$ and $E_{ijk} \in N(0, \sigma_{rep}^2)$. In contrast to $\sigma_{area,agg}^2$, $\sigma_{area,hi}^2$ only accounts for the medium scale variation in the blend, i.e. the variation between the areas within a layer.

The model is a hierarchical (also called nested or stratified) model. A detailed description of nested models is also given in [19].

When there is no difference between layers (i.e. $\sigma_{layer}^2 = 0$) the aggregated model (1) and the hierarchical model (2) are identical, i.e. $\sigma_{area,agg}^2 = \sigma_{area,hi}^2 = \sigma_{area}^2$. The hierarchical model corresponds to the definition of homogeneity used in the cases in Appendix F.

The hierarchical model including differing small scale variation

In practice when sampling replicates from the same area the replicates are sampled as close as possible without overlapping. The reason is to avoid spots with a high degree of deblending caused by the insertion of the thief when collecting previous samples. Thus the variation between replicates includes variation between content in neighbouring spots within an area, i.e. variation in the batch measured on a very small scale.

In a situation where deblending has occurred in one layer the variation between the content in neighbouring spots within an area may differ from the corresponding variation in other layers, hence the small scale variation may depend on the layer.

Further, the static pressure in the blend is lower near the surface than near the bottom. This difference in static pressure may affect the sampling results (see e.g. Berman [2]). One example would be the difference in static pressure causing the variation between the content in replicate samples being larger when sampled from the bottom layer than when samples from the top layer (see e.g. Appendix F).

The (simulated) variation between replicate samples is seen to be the result of several causes: small scale variation in the blend and variation due to sampling and chemical analysis. At least two physical causes (small scale variation and variation due to sampling) may lead to a variation between replicates, σ_{rep}^2 , that depends on the layer.

Neither the aggregated model (1) nor the hierarchical model (2) accounts for this type of variation. However, with the following extension to the hierarchical model (2), small scale variation is accounted for:

$$E_{ijk} \in N(0, \sigma_{rep,i}^2), \quad \begin{cases} i = 1, 2, 3 \\ j = 1, 2, 3, 4 \\ k = 1, 2, 3 \end{cases} \quad (3)$$

2.3 Simulated samples

The full hierarchical model (the hierarchical model (2) with the extension (3)) is able to account for three different types of variation in the batch: variation measured on a large scale (the variation between layers), a medium scale (variation between areas within a layer) and the small scale variation (variation between replicates within an area). This ability makes it a very versatile model. The drawback is that the model includes six parameters (μ , σ_{layer}^2 , $\sigma_{area,hi}^2$, $\sigma_{rep,top}^2$, $\sigma_{rep,middle}^2$ and $\sigma_{rep,bottom}^2$), hence it is very difficult in a clear way to analyse and present results with all parameters varying at a time. Therefore in each series of simulated batches some of the parameters are kept fixed depending on the aim (hypothesis) of the model under analysis.

Thus, the full hierarchical model (model (2) with the extension (3)) was used to generate a number of datasets all with a total of 36 samples equally distributed over 3 layers, 4 areas within each layer and 3 replicates within each area. The sampling plan is shown in the first three columns of Table 1, page 52. On the whole the number of samples corresponds to the number of samples in Appendix F as well as the sampling plan suggested by PQRI [18]. In each analysis 500 batches are simulated for each combination of the parameter values.

In the analysis the sample content is measured in percent of label claim, LC. Thus $\mu = 100$ corresponds to LC. However, as the overall mean content in the batch is balanced out in the calculations the results are independent of the value of the overall mean content. Further the results do only depend on the ratios $\sigma_{layer}/\sigma_{rep}$ and $\sigma_{area,hi}/\sigma_{rep}$ (or $\sigma_{area,agg}/\sigma_{rep}$), and therefore the value of the variance components is measured relatively to σ_{rep}^2 (or $\sigma_{rep,top}^2$ in a model with extension (3)), i.e. either $\sigma_{rep} = 1$ or $\sigma_{rep,top} = 1$.

Two types of analyses are made in the following sections. The first type of analysis is described in Section 3. It focuses on factors influencing the mean content of a sample corresponding to the terms A_i , $A(L)_{j(i)}$ and L_i in the

aggregated model (1) and the hierarchical model (2). This corresponds to an analysis of the large and the medium scale variation. The other type of analysis, described in Section 4 focuses on factors influencing the term E_{ijk} in extension (3) of the full hierarchical model. Thus this is an analysis of the small scale variation.

3 Assessing factors with influence on the mean content of the active component in a sample

In Section 2 it is described how a number of samples were simulated from a hierarchical model. These simulated samples represent real samples and real batches.

In practice a statistical method corresponding to the aggregated model (1) is (most often) used to analyse blend sample results (see eg. [22]). In essence the acceptance criteria suggested by PQRI [18] is also based on this model, i.e. the acceptance criteria does not explicitly take into account a possible large scale variation between layers.

In this section the simulated batches (some of which contain variation between layers) are analysed in accordance with both the aggregated model (1) and the hierarchical model (2). The robustness and the power of these methods is assessed as a function of the variation between layers, σ_{layer} , and the variation between areas within a layer, $\sigma_{area,hi}$.

Finally the effect of including an external factor (sampling thieves) in the analyses of the mean content of the collected samples is discussed.

The tests are conducted using the GLM procedure in SAS. In simulations of samples for the analysis $\mu = 100$ and $\sigma_{rep}^2 = 1$.

3.1 Large and medium scale homogeneity assessed according to the aggregated model (1)

A definition of blend homogeneity related to the aggregated model (1) is the situation where there is no variation between areas, i.e. $\sigma_{area,agg}^2 = 0$. A statistical test of the hypothesis $\sigma_{area,agg}^2 = 0$ is described e.g. by Montgomery [19]. The test statistic is

$$Z_{area,agg} = \frac{S_{area,agg}^2}{S_{rep}^2}, \quad (4)$$

where

$$S_{area,agg}^2 = \frac{3 \times \left[\sum_{i=1}^{12} (\bar{y}_i - \bar{y}_{..})^2 \right]}{12 - 1} \quad (5)$$

and

$$S_{rep}^2 = \frac{\sum_{i=1}^{12} \sum_{j=1}^3 (y_{ij} - \bar{y}_i)^2}{12 \times (3 - 1)}. \quad (6)$$

\bar{y}_i is the mean of the three replicates in area i and $\bar{y}_{..}$ is the overall mean.

Under the aggregated model with $\sigma_{area,agg}^2 = 0$ the test statistic follows a $F(12 - 1, 12 \times (3 - 1))$ distribution, and therefore $\sigma_{area,agg}^2$ is declared significantly different from 0 at the 5 % level when $Z_{area,agg} > F(11, 24)_{.95} = 2.18$. Under the aggregated model (1)

$$\frac{S_{area,agg}^2 / (3 \times \sigma_{area,agg}^2 + \sigma_{rep}^2)}{S_{rep}^2 / (\sigma_{rep}^2)} \in F(12 - 1, 12 \times (3 - 1)). \quad (7)$$

As the samples were simulated from a model including the variation between layers, the samples from the same layer are centered around the mean content in that layer. Hence the averages \bar{y}_i s are correlated, and therefore the underlying assumption from the aggregated model (1), that the mean content in the areas are uncorrelated, is violated. Strictly spoken this means that $\sigma_{area,agg}^2$ from model (1) does not exist in a context with σ_{layer}^2 and $\sigma_{area,hi}^2$.

Under the hierarchical model (2) - which corresponds to the model from which the samples are simulated - the overall variation between areas is separated into

the variation between layers, σ_{layer}^2 , and the variation between areas within a layer, $\sigma_{area,hi}^2$. In relation to the specific sampling plan from which the samples are sampled (i.e. four areas within each layer) the mean value of $S_{area,agg}^2$ is $E[S_{area,agg}^2] = \frac{24}{11}\sigma_{layer}^2 + \frac{33}{11}\sigma_{area,hi}^2 + \sigma_{rep}^2$.

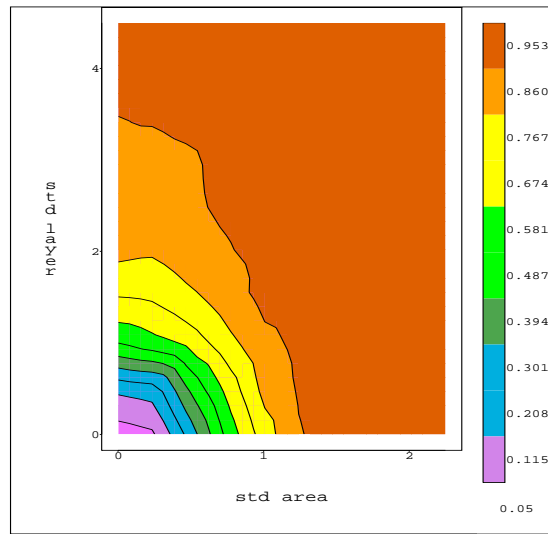


Fig. 3: The probability of declaring the variation between areas under the aggregated model, $\sigma_{area,agg}^2$, significant at the 5% level as a function of $\sigma_{area,hi}$ and σ_{layer} . This is also called the power of the test. The probability is low when both $\sigma_{area,hi}$ and σ_{layer} are small. When $\sigma_{area,hi}$ is larger than $1.25 \times \sigma_{rep}$ the probability is more than 0.95 no matter the value of σ_{layer} . Similarly when σ_{layer} is larger than $3.5 \times \sigma_{rep}$ the probability of testing $\sigma_{area,agg}$ significant is more than 0.95, irrespective of the value of $\sigma_{area,hi}$

This is also seen from Figure 3. The figure shows level curves for the power of the test in (7) as a function of $\sigma_{area,hi}$ and σ_{layer} . The power of the test is the probability of declaring $\sigma_{area,agg}^2$ significant (here at the 5% level). The probability is found for each set of parameter values by testing simulated samples from 500 batches.

From Figure 3 it is seen that for fixed values of σ_{rep}^2 the level curves correspond to ellipses $\sqrt{\frac{24}{11}\sigma_{layer}^2 + \frac{33}{11}\sigma_{area,hi}^2} = \text{constant}$.

More explicit it is seen that, when $\sigma_{area,hi}$ is larger than $1.25 \times \sigma_{rep}$ the probability of finding $\sigma_{area,agg}^2 > 0$ is at least 0.95. When $\sigma_{area,hi}$ is smaller the probability of declaring $\sigma_{area,agg}^2$ significant depends on the value of σ_{layer} . The plot shows that because the variation between layers is not specifically accounted for in the aggregated model (1), variation between layers in the batch is identified as variation between areas. When the variation between layers, σ_{layer} , is more than $3.5 \times \sigma_{rep}$ the probability of testing $\sigma_{area,agg}^2 > 0$ is at least 0.95. In conclusion the test in 7 really measures $\frac{24}{11}\sigma_{layer}^2 + \frac{33}{11}\sigma_{area,hi}^2 + \sigma_{rep}^2$ relatively to σ_{rep}^2 .

3.2 Large and medium scale homogeneity assessed according to the hierarchical model (2)

A definition of blend homogeneity related to the hierarchical model (2) is $\sigma_{layer}^2 = 0$ and $\sigma_{area,hi}^2 = 0$. The test statistic for the two hypotheses $\sigma_{layer}^2 = 0$ and $\sigma_{area,hi}^2 = 0$ are respectively (see e.g. [19])

$$Z_{layer} = \frac{S_{layer}^2}{S_{area,hi}^2} \quad (8)$$

and

$$Z_{area,hi} = \frac{S_{area,hi}^2}{S_{rep}^2}, \quad (9)$$

where

$$S_{layer}^2 = \frac{4 \times 3 \times \left[\sum_{i=1}^3 (\bar{y}_{i..} - \bar{y}_{...})^2 \right]}{3 - 1} \quad (10)$$

and

$$S_{area,hi}^2 = \frac{3 \times \left[\sum_{i=1}^3 \sum_{j=1}^4 (\bar{y}_{ij.} - \bar{y}_{i..})^2 \right]}{3 \times (4 - 1)} \quad (11)$$

and

$$S_{rep}^2 = \frac{\sum_{i=1}^3 \sum_{j=1}^4 \sum_{k=1}^3 (\bar{y}_{ijk} - \bar{y}_{ij.})^2}{3 \times 4 \times (3 - 1)} \quad (12)$$

$\bar{y}_{...}$ is the overall mean, $\bar{y}_{i..}$ is the mean of the i 'th layer and $\bar{y}_{ij.}$ is the mean of the j 'th area in the i 'th layer.

The critical intervals for the test statistics are $Z_{layer} > F(2, 9)_{.95} = 4.26$ and $Z_{area,hi} > F(9, 24)_{.95} = 2.30$

Under the hierarchical model the variation between areas, $\sigma_{area,hi}^2$, is corrected for the contribution from layers, σ_{layer}^2 . Hence, in this case the test of the variation between areas, $\sigma_{area,hi}^2 = 0$, is independent of the value of σ_{layer}^2 .

However, the test statistic, Z_{layer} , depends on $S_{area,hi}^2$ and therefore the test of σ_{layer}^2 is expected to depend on the value of $\sigma_{area,hi}^2$.

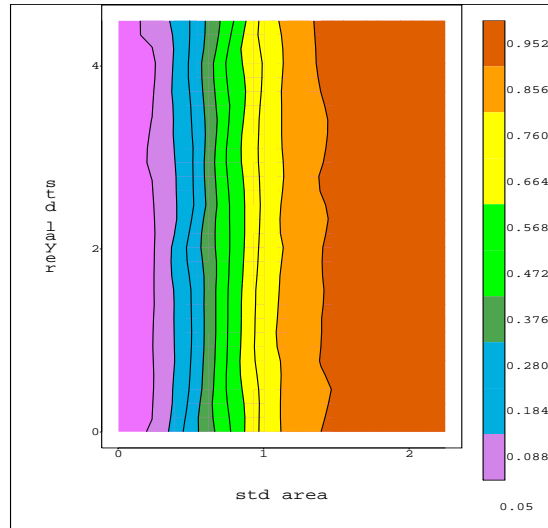


Fig. 4: The power of the 5% level test of $\sigma_{area,hi}^2$ in the hierarchical model (2). When $\sigma_{area,hi}$ is approximately $1.5 \times \sigma_{rep}$ the probability of declaring significance is more than 0.95. The probability is independent of σ_{layer} .

For the test at the 5 % level Figure 4 shows the power of the test of $\sigma_{area,hi}^2$ as a function of $\sigma_{area,hi}$ and σ_{layer} .

As expected it is seen that now that the variation between layers is accounted for in the model the test of the variation between areas does no longer depend of the value of σ_{layer} . In agreement with Figure 3, Figure 4 shows that when $\sigma_{area,hi}$ is approximately $1.5 \times \sigma_{rep}$, the probability of declaring $\sigma_{area,hi}^2$ significant is at least 0.95.

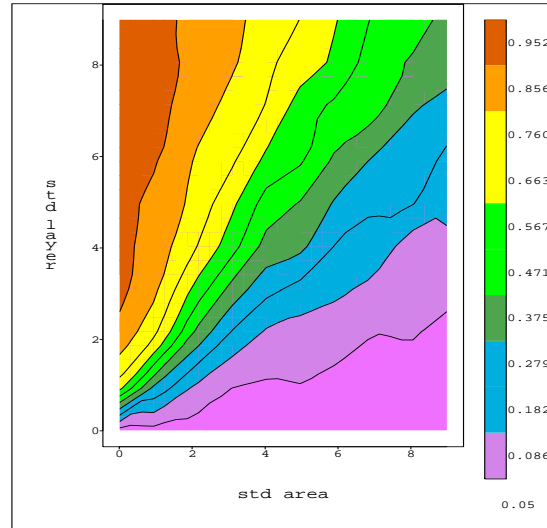


Fig. 5: The probability of the 5% level test of σ_{layer}^2 in the hierarchical model (2). The power is dependent of σ_{layer} and $\sigma_{area,hi}$.

Figure 5 shows the power of the test of σ_{layer}^2 as a function of $\sigma_{area,hi}$ and σ_{layer} . As expected the power of the test of σ_{layer}^2 depends on the value of $\sigma_{area,hi}$, namely in essence of $\frac{\sigma_{layer}^2}{1+3 \times \sigma_{area,hi}^2}$.

If the factors describing the variation between layers had been simulated with three fixed (predefined) levels the test statistic would follow a non-central F-distribution, but the overall picture would not be changed.

3.3 The effect of including sampling thieves in the model

So far the homogeneity of the blend has been modelled by models taking into account the large and the medium scale variation. Blend samples results are however not entirely a consequence of the batch homogeneity. Sometimes it is relevant to investigate whether other factors such as sampling thief or sampling technique have an effect on the content in a sample. This is the situation in Appendix F. Also Berman [23] has assessed the significance of factors as

sampling thief and technique using an analysis of variance.

A number of batches in which two different thieves were used for sampling were simulated. The experimental design is shown in Table 1. The model is

$$Y_{ijkl} = \mu + L_i + A(L)_{j(i)} + t_k + E_{ijkl}, \quad \begin{cases} i = 1, 2, 3 \\ j = 1, 2, 3, 4 \\ k = 1, 2 \\ l = 1, 2, 3 \end{cases}, \quad (13)$$

where t_k is the effect of thief k , and $\sum t_k = 0$. Apart from that the assumptions are identical to those from the hierarchical model (2).

For practical reasons the following parameters are fixed: $\sigma_{layer} = 2$, $\sigma_{rep} = 1$ and $\mu = 100$.

The test statistic for the hypothesis $t_k = 0$, $\sigma_{layer}^2 = 0$ and $\sigma_{area,hi}^2 = 0$ are now

$$Z_{thief} = \frac{S_{thief}^2}{S_{rep}^2}, \quad (14)$$

$$Z_{layer} = \frac{S_{layer}^2}{S_{area,hi}^2}, \quad (15)$$

and

$$Z_{area,hi} = \frac{S_{area,hi}^2}{S_{rep}^2}. \quad (16)$$

S_{rep}^2 denotes the estimate of the variance between replicates, σ_{rep}^2 , obtained from the residual sum of squares corresponding to model (13), $S_{area,hi}^2$ denotes the mean sum of squares corresponding to variation between areas, corrected for layers and thieves in analogy with (11), S_{layer}^2 in analogy with (10) denotes the mean sum of squares corresponding to the variation between layers. Finally

No.	Layer	Area	Thief
1	top	A	X
2	top	A	X
3	top	A	Y
4	top	B	X
5	top	B	Y
6	top	B	Y
7	top	C	X
8	top	C	X
9	top	C	Y
10	top	D	X
11	top	D	Y
12	top	D	Y
13	middle	A	X
14	middle	A	X
15	middle	A	Y
16	middle	B	X
17	middle	B	Y
18	middle	B	Y
19	middle	C	X
20	middle	C	X
21	middle	C	Y
22	middle	D	X
23	middle	D	Y
24	middle	D	Y
25	bottom	A	X
26	bottom	A	X
27	bottom	A	Y
28	bottom	B	X
29	bottom	B	Y
30	bottom	B	Y
31	bottom	C	X
32	bottom	C	X
33	bottom	C	Y
34	bottom	D	X
35	bottom	D	Y
36	bottom	D	Y

Table 1: All simulated batches include 36 samples: 12 samples from each of three layers and 3 samples from each of four areas in each layer. Some batches were simulated with two different thieves included in the experimental design as shown in the last column. In the simulation of other batches the variation between replicate samples, σ_{rep}^2 , depends on the layer from which the samples are sampled.

S_{thief}^2 denotes the mean sum of squares corresponding to variation between thieves.

The critical value for the test statistic corresponding to thieves is $Z_{thief} > F(1, 23)_{.95} = 4.3$. The test statistic corresponding to layers and areas within a layer are unchanged from the hierarchical model (1).

As there were only three replicate samples in each area, the use of two different thieves could not be completely balanced within an area, and therefore the partitioning of the variation is not straightforward. However, as the design allows for an estimation of the variation between replicates and as the use of the two thieves has been balanced over areas and layers, it is seen from (14) that the test of the effect of sampling thief, t_k , is independent of σ_{layer} and $\sigma_{area,hi}$.

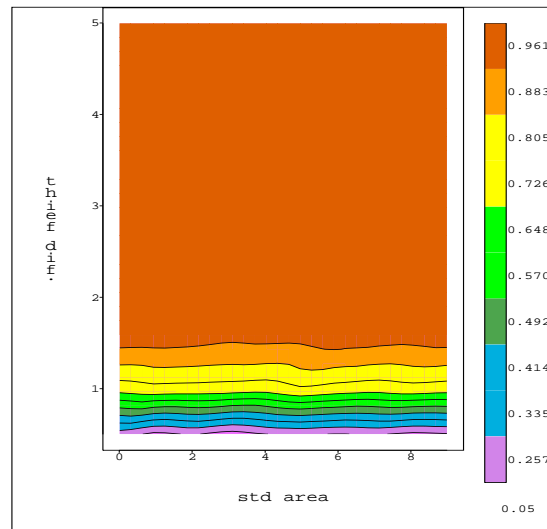


Fig. 6: The power of the 5% level test of the thief effect, t_k , i.e. the probability of declaring the thief effect significant.

Figure 6 illustrates the power of the 5 % level test of t_k . The power is plotted as a function of $\sigma_{area,hi}^2$ and the effect of the thieves. The effect is the difference between the mean of samples collected with one thief and the mean of samples collected with the other thief.

The power of the test of the thief effect is independent of $\sigma_{area,hi}$ and a difference between samples collected with each of the two thieves of larger than 1.5% of LC has a probability of at least 0.95 of being detected with the given design.

From expression (16) the test of $\sigma_{area,hi}^2$ is expected to be independent of both σ_{layer} and t_k . Expression (15) shows that the power of the test of σ_{layer}^2 is independent of t_k but not of σ_{area}^2 . Hence, plot of the power of these two tests are similar to Figure 4 and Figure 5.

3.4 Conclusion

Samples from a number of batches have been simulated according to a hierarchical model that explicitly takes into account respectively the large and the medium scale variation in the batch. Further, it is assumed that the variation between replicate samples is independent of the layer and that there is no interaction between the factors in the model.

Under the assumption that the homogeneity of real batches in this way may consist of both a large and a medium scale variation the power of two statistical methods to assess batch homogeneity has been found.

The first statistical method (the aggregated model) corresponds to an 'aggregated' definition of homogeneity in the sense that large and medium scale variation in the batch is assessed as a whole.

The other statistical method (the hierarchical model) corresponds to a homogeneity definition with two different criteria; one explicitly regarding the large scale variation and the other explicitly regarding the medium scale variation.

The probability of declaring inhomogeneity is basically the same for analysis corresponding to each of the two model. The most important difference between the two types of analysis is that when inhomogeneity is declared according to the aggregated analysis the result does not reveal whether this inhomogeneity is due to large, medium or small scale variation in the batch. However, the hierarchical model explicitly assesses respectively large and medium scale variation.

The power of the respective tests are shown in Figure 3 to 5. It should be noted that σ_{layer} and $\sigma_{area,hi}$ are measured relatively to σ_{rep} .

Finally, for the given design the power of the test of a thief effect was assessed in Figure 6. In these simulations the variation between layers were held fixed ($\sigma_{layer} = 2$). However, the test of the thief effect is independent of σ_{layer} and $\sigma_{area,hi}$.

With the given design a difference greater than 1.5 in mean content in samples from the two thieves has a probability of at least 0.95 of being detected.

4 Assessing factors with influence on the variation between replicates

In Section 3 the robustness and power of statistical methods to evaluate the large scale and medium scale variation in the blend were assessed by means of a General Linear Model. In this section the robustness and power of a statistical method to evaluate the small scale variation is assessed. Small scale variation is variation in the blend measured on a small scale, i.e. variation between neighbouring samples.

However in practice, variation between neighbouring samples includes both small scale variation in the blend as well as sampling error and variation due to the chemical analysis. In practice replicate samples from exactly the same spot is avoided because of the risk of subsequent samples being biased due to deblending from taking the first sample. Therefore neighbouring samples are also referred to as replicate samples.

The method presented in this section can be used to assess factors with influence on either the small scale variation or eg. the size of the variation due to the sampling, depending on the experimental design and the assumptions made.

In this analysis the factor under consideration is the effect of layers: Layers having an effect on the small scale variation means the special kind of inhomogeneity where for example the variation between neighbouring spots within an area is smaller in the top layer, $\sigma_{rep,top}^2$, than in the bottom layer, $\sigma_{rep,bottom}^2$.

A tendency like this was found in Appendix F.

However, differences between the variation between replicates in the top layer, $\sigma_{rep,top}^2$, and the variation between replicates in the bottom layer, $\sigma_{rep,bottom}^2$, may also be due to the fact that it is more difficult to handle a long thief in the bottom of the blend than handling a short thief in the top layer. Also the static pressure in the bottom of a batch is greater than near surface, which may result in different types of sampling bias [2] for example differences in the variation between replicates. These are examples of layers having an effect on the size of the sampling error (variation).

In this analysis it is assumed that the variation between replicates represents the small scale variation. The variation due to the chemical analysis is assumed to be neglectable.

For the analysis of the small scale variation samples from a number of batches were simulated according to the hierarchical model (2) with extension (3).

In the simulations $\sigma_{rep,top} = 1$. As one set of replicates is sampled from each of four areas within each of three layers, the design is balanced. See also Table 1.

For each area,

$$S_{rep,ij}^2 = \frac{\sum_{k=1}^3 (y_{ijk} - \bar{y}_{ij.})^2}{3 - 1}, \quad \begin{cases} i = 1, 2, 3 \\ j = 1, 2, 3, 4 \\ k = 1, 2, 3 \end{cases} \quad (17)$$

provides an estimate of the variance between replicates in that area.

Under the assumption of a normal distribution of replicates, (3), it follows that

$$S_{rep,ij}^2 \in \sigma_{rep,i}^2 \chi^2(2)/2, \quad (18)$$

see e.g. Montgomery [19], and therefore standard methods for analysis of linear models for normally distributed observations are not applicable to assess hypotheses concerning $\sigma_{rep,i}^2$.

Bartlett and Kendall [24] have suggested approximate tests for the analysis of variance heterogeneity. However, utilizing the relation between the χ^2 -distribution and the gamma-distribution, models for variance heterogeneity may be investigated using the theory for the so-called generalized linear models [25].

The hypothesis

$$\sigma_{rep,top}^2 = \sigma_{rep,middle}^2 = \sigma_{rep,bottom}^2 \quad (19)$$

may be assessed e.g. by means of the GENMOD procedure of SAS specifying the model

$$\ln(\sigma_{rep,i}^2) = \mu^* + \alpha_i, \quad (20)$$

with $\alpha_{top} = 0$ and testing the hypothesis

$$\alpha_{middle}^* = \alpha_{bottom}^* = 0. \quad (21)$$

In the simulation a test at the 5 % level was performed.

Figure 7 shows the level curves for the power of the test as function of $\ln(\sigma_{rep,middle})$ and $\ln(\sigma_{rep,bottom})$. It is seen that the level curves have ellipsoid shape. The reason is that the power depends on

$$\begin{aligned} & \left[\frac{\ln(\sigma_{rep,middle}^2) + \ln(\sigma_{rep,bottom}^2)}{2} \right]^2 + 3 \left[\frac{\ln(\sigma_{rep,middle}^2) - \ln(\sigma_{rep,bottom}^2)}{2} \right]^2 \\ & = (\ln(\sigma_{rep,middle}) + \ln(\sigma_{rep,bottom}))^2 + 3 [\ln(\sigma_{rep,middle}) - \ln(\sigma_{rep,bottom})]^2 \\ & = C. \end{aligned} \quad (22)$$

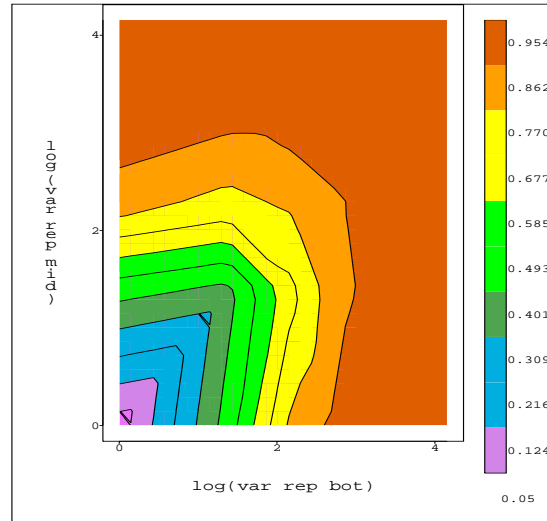


Fig. 7: Level curves for the power of the 5% level test for a dependency of σ_{rep}^2 on layers. The power is plotted against $\sigma_{rep,middle}$ and $\sigma_{rep,bottom}$.

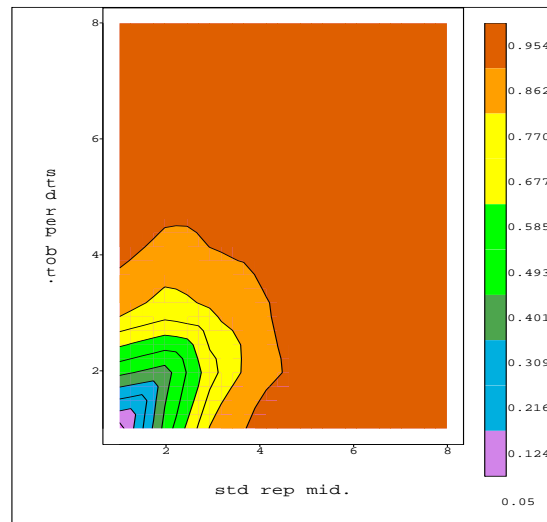


Fig. 8: Level curves for the power of the 5% level test for a dependency of σ_{rep}^2 on layers plotted against $\sigma_{rep,middle}$ and $\sigma_{rep,bottom}$.

Furthermore Figure 8 shows the probability of declaring a significant difference between $\sigma_{rep,top}^2$, $\sigma_{rep,middle}^2$ and $\sigma_{rep,bottom}^2$ as a function of $\sigma_{rep,middle}$ and $\sigma_{rep,bottom}$. When at least one of $\sigma_{rep,middle}$ and $\sigma_{rep,bottom}$ is more than 4.5 times as large as $\sigma_{rep,top}$ the probability of declaring significance is more than 0.95.

The shape of the level curves is similar for other values of $\sigma_{rep,top}$.

4.1 Conclusion

In this section the theory of generalized linear models is presented as a tool to analyse e.g. the influence of layers on the small scale variation.

For samples simulated from model (2) with the extension (3) it has been shown that for the 5 % level test the variation between replicates $\sigma_{rep,i}$ has to be 4.5 times larger in one layer than the variation between replicates in another layer for the effect of layers to be declared significant with a probability of at least 0.95.

Depending on the experimental design and the assumptions made the method presented in this section can also be used to assess factors influencing sample error (variation). For example the method can be used to test if one sampling thief leads to larger variation between replicate samples than another thief.

5 Conclusion

In this appendix a statistical model describing blend homogeneity is presented. The model is a hierarchical model explicitly taking into account both large, medium and small scale variation in the blend. Large scale variation is variation measured on a large scale, e.g., top, middle and bottom layer of the blend. Medium scale variation is in this model variation between areas within a layer. Small scale variation is variation between the mean content in the neighbourhood samples within an area.

As both large, medium and small scale variation may occur in real blend

batches all batches and samples were simulated from this hierarchical model.

However, the analysis of the simulated samples were conducted both in accordance with an aggregated model that as a measure of inhomogeneity uses an aggregate of the medium and the large scale variation, and in accordance with a hierarchical model explicitly taking into account the large and the medium scale variation. In essence the aggregated model relates to the acceptance criteria suggested by FDA [11] and PQRI [18].

The power of the tests of the homogeneity according to respectively the aggregated and the hierarchical model is found. Basically the probability of declaring inhomogeneity is similar for analysis corresponding to each of the two models. The power of the respective tests are given in more details in Figure 3 to Figure 5.

As the probability of declaring inhomogeneity is basically the same for analysis corresponding to each of the two models, the aggregated model could be useful if the purpose of the analysis is to detect inhomogeneity and there is no special reason to suspect large scale inhomogeneity (i.e. variation between layers).

However, by using the hierarchical analysis explicit knowledge of respectively the large and the medium scale variation in the batch is provided. Hence, in a situation with suspicion of large scale inhomogeneity or if the purpose is to get more specific knowledge about the homogeneity of the batch, an analysis in accordance with the hierarchical model should be chosen.

The aggregated as well as the hierarchical model describe inherent variations in the batch. By an example it has been shown how external factors such as thieves can be included in the hierarchical model. Other external factors as sampling technique or sampling personnel could be included in the model as well (see Appendix F).

In the specific design used in this appendix (with $\sigma_{rep}^2 = 1$) the 5% level test for the thief effect has a probability of at least 0.95 of declaring significance when the difference between the mean content in samples collected with each of the two thieves is larger than 1.5.

The method of Generalized Linear Models was introduced to conduct a sep-

arate analysis of the variation between replicate samples. It was shown that when the variation between replicates is more than 4.5 times as large as the variation between replicates in another layer, probability of detecting an effect of layers is at least 0.95.

For all tests in this appendix it should be noted that the power of the tests corresponds to the specific experimental plan used. For other experimental plans the power of the corresponding test can be found after same principles as in this appendix.

Further, when using the methods to assess factors with influence on the mean content (see Section 3) it is assumed that the variation between replicates is the same. In the analysis in this appendix only batches with constant variation between replicates were simulated for this analysis. However, when analysing the mean content in samples from real batches there is a risk that the variation between replicates is not constant. The method presented in Section 4 can be used to verify this assumption. Therefore in practice a suitable procedure is to use the methods presented in this appendix in reversed order, i.e. first to analyse the variation between replicates and if this variation is constant, then continue with the analysis of the mean content. If the variation between replicates is not the same, a so-called weighted squares of means method should be used (see e.g. [19]).

B

Paper B

**Comprehensive measures of
blend uniformity**

1 Introduction

An important quality characteristic in tablet production is the homogeneity of the tablets, i.e. the variation in content of the active ingredient. As this homogeneity is closely related to the homogeneity of the blend, the homogeneity is also an important quality measure of the blend.

One measure of homogeneity is defined in Appendix A. Here homogeneity is defined as a relevant component of variance being equal to zero. However, ultimate homogeneity like this is neither necessary nor practical obtainable - neither for the tablets nor for the blend. Therefore the definition of homogeneity in Appendix A is not suitable as acceptance criteria. Instead the quality of the blend (or the product) is evaluated relatively to practical criteria assessing if the homogeneity of the blend (product) is satisfactory for the blend (product) to serve its purpose. Satisfactory homogeneity is here referred to as uniformity.

The issue is then to find an acceptance criteria corresponding to satisfactory homogeneity/uniformity.

In Appendix A it is shown how to model homogeneity in the blend using three different scales of homogeneity: large, medium and small scale variation. For exploratory investigations of the blending process, an analysis based on these three scales of homogeneity is relevant. However, both existing and suggested acceptance criteria for blend uniformity are based on the total variation in the blend, for example through the sample coefficient of variation (see e.g. [2] and [22]). The aim of the discussion in this appendix is to investigate how the total variation in a blend can be interpreted in case of different types of inhomogeneity.

The discussion is exemplified by the first stage of a procedure for blend validation suggested by PQRI [18]. However, the principles in the discussion of interpreting the total variation in the blend would be the same had a different acceptance criteria been chosen, e.g. the acceptance criteria recommended by FDA [11] or the standard prediction interval (SDPI) method suggested by PDA [22].

In the first stage of the procedure suggested by PQRI it is recommended to assay one sample per area (location) in the batch. Samples should be collected from at least ten different areas. The relative standard deviation (sample stan-

standard deviation divided by the average) for the samples should be less than or equal to 5.0% and all individual results should be within 90.0% - 110.0 % of the mean result (of all samples). If these two criteria are met the first stage of the suggested procedure is passed.

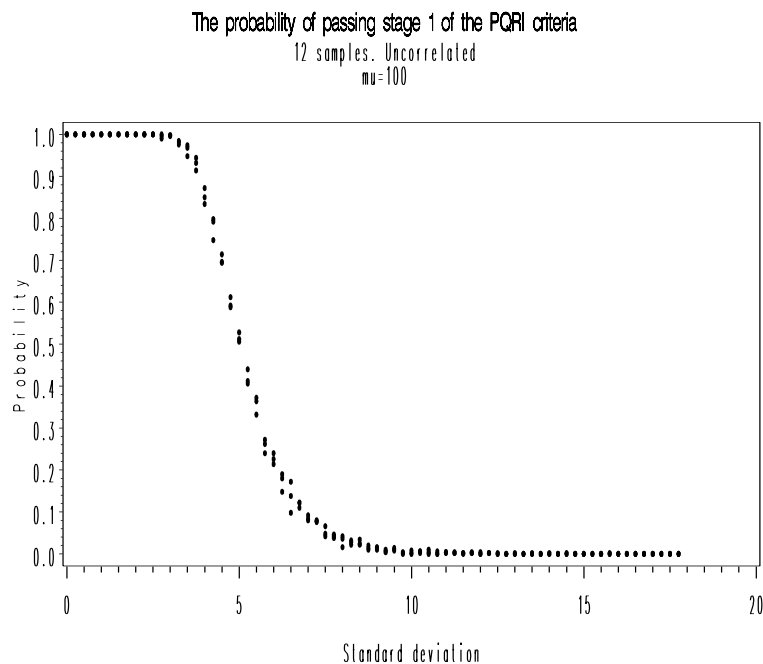


Fig. 1: The probability of passing the first stage of the PQRI acceptance criteria for blend uniformity validation [18]. The probability is plotted as a function of the total standard deviation in the blend.

The properties (OC-surface) of this acceptance criteria can be investigated for example as a function of the mean content and the standard deviation in the batch (see e.g. Appendix D and Appendix C). However, in Figure 1 the probability of acceptance is plotted as a function of the standard deviation in the batch. The batch mean content is fixed at label claim, LC ¹. The OC-surface correspond to a sampling plan where one sample is collected from each of 12 areas in the blend. The plot is based on simulations under the assumption that the 12 samples are uncorrelated and identically (normally) distributed. The question regarding the samples being uncorrelated will be discussed later in

¹In this appendix all measurements are given in percent of LC. Hence, for a batch with mean content equal to label claim, $\mu = 100$

this appendix.

Besides a slight variation due to the simulation procedure the plot shows a slim, smooth curve indicating that in the situation of uncorrelated, identically (normally) distributed samples, there exists an unambiguous relation between the uniformity (total variation) of the blend and the probability for the batch to pass the acceptance criteria. The question is how this relation is affected by the interpretation of the total variation in the blend.

In the analysis in this appendix it should be noted that the values of the variance components are no longer measured relatively to the small scale variation as in Appendix A. All variance components are measured absolutely. For practical purposes the small scale variation (variation between replicates from the same area) has been fixed, $\sigma_{rep}^2 = 1$, in these calculations and simulations. The results in this appendix are only valid for the absolute value of the variance components on which the calculations are based.

The sampling plan is seen in Table 1.

No.	Layer	Area
1	top	A
2	top	B
3	top	C
4	top	D
5	middle	A
6	middle	B
7	middle	C
8	middle	D
9	bottom	A
10	bottom	B
11	bottom	C
12	bottom	D

Table 1: Most of the simulations in this appendix are based on 12 samples from each batch. The distribution of the samples in the batch appears from the table.

2 Batches with medium scale variation (variation between areas)

As mentioned above the first stage of the procedure suggested by PQRI [18] recommends to assay one sample per area in the batch. Samples should be collected from at least ten different areas. In the simulations discussed in this appendix samples from twelve different areas were collected. This sampling plan does not specifically refer to layers in the batch. Therefore, as a first example batches with no variation between layers and a fixed standard deviation corresponding to replicates, $\sigma_{rep} = 1$, are simulated. With no variation between layers the hierarchical model (2) in Appendix A from which the batches are simulated, is identical to the aggregated model (1) in Appendix A, hence, $\sigma_{area,hi}^2 = \sigma_{area,agg}^2 = \sigma_{area}^2$.

Two different principles of determining the total variation in the batch is presented in the following.

2.1 The total variation from the ANOVA table

One way of defining the total variation in the batch is to use the expression for the total variance obtained from the analysis of variance table corresponding to an analysis of sample results obtained from the specific sampling plan, in this case the plan sketched in Table 1. The model for this analysis is

$$Y_{ij} = \mu + A_i + E_i, \quad i = 1, \dots, 12. \quad (1)$$

Y_{ij} is the content of a sample, μ the mean content in the batch, A_i the effect corresponding to the area from which the sample is taken, and E_i is the error term also including the variation due to sampling. As only one sample from each area is collected the terms A_i and E_i have identical indices. Further $A_i \in N(0, \sigma_{area,agg}^2)$ and $E_i \in N(0, \sigma_{rep}^2)$. The analysis of variance table corresponding to this model is given in Montgomery [19].

In this case where one sample is collected from each of twelve areas the analysis of variance (ANOVA) table is

ANOVA Table			
Source of variation	Sums of squares	Degrees of freedom (df)	Expected mean square (EMS)
Between areas	$SS_{area,agg}$	12 - 1	$\sigma_{area,agg}^2 + \sigma_{rep}^2$
Error	SS_{rep}	0	σ_{rep}^2
Total	SS_{total}	12 - 1	

Table 2: The analysis of variance table corresponding to model (1).

As the terms in the 'sums of squares' column are additive the total variation in the blend, σ_{total}^2 , assessed under this scheme is estimated from

$$\begin{aligned}
 \sigma_{total}^2 &= \frac{SS_{area,agg} + SS_{rep}}{df_{total}} \\
 &= \frac{df_{area,agg} \times MSE_{area,agg} + df_{rep} \times MSE_{rep}}{df_{total}} \\
 &= \frac{11 \times (\sigma_{area,agg}^2 + \sigma_{rep}^2)}{11} \\
 &= \sigma_{area,agg}^2 + \sigma_{rep}^2. \quad (2)
 \end{aligned}$$

2.2 The variance on a randomly sampled unit from the batch

Another definition of the total variation in the blend is to consider the variance on a randomly sampled unit. In relation to model (1) this variance is $\sigma_{sample}^2 = \sigma_{area,agg}^2 + \sigma_{rep}^2$. In this situation with one replicate per area and no variation between layers, the total variation in the blend determined from the ANOVA table is identical to the variance on a randomly sampled unit from the blend.

Figure 2 shows the level curves for the OC-surface of the PQRI acceptance criteria as a function of the mean content and the total standard deviation in the blend, $\sigma_{sample} = \sigma_{total}$, determined from the expression (2).

The figure shows that when the total standard deviation is less than approximately 3%, the probability of passing the first stage of the PQRI criteria is at least 0.95. Further, it should be noted that even batches with a mean content relatively far from LC ($\mu = 100$) have a relative high probability of accep-

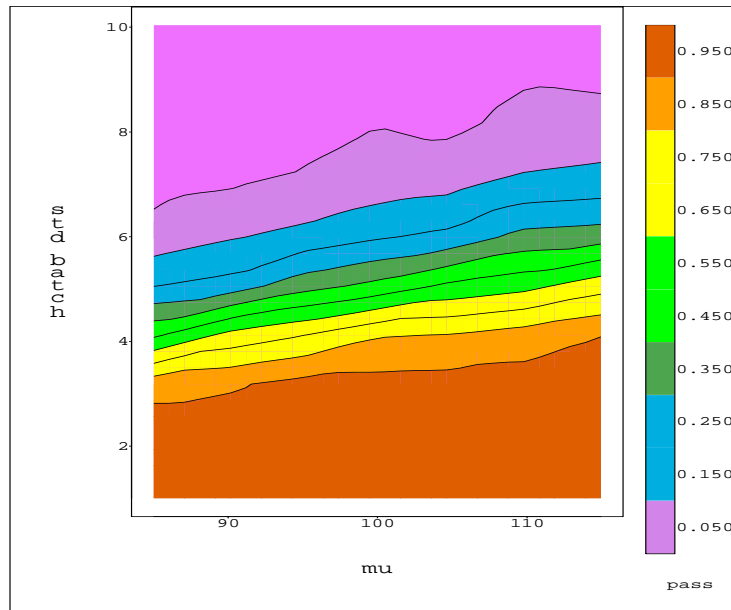


Fig. 2: Level curves for properties (OC-surface) of the first stage of the procedure for blend validation suggested by PQRI [18]. The OC-surface is plotted against the mean content in the blend, μ , and the total standard deviation in the blend in a situation with no variation between layers, $\sigma_{total} = \sqrt{\sigma_{area,agg}^2 + \sigma_{rep}^2}$. Under model (1) the expression for the total standard deviation determined from the ANOVA table is identical to the expression for the total standard deviation on a randomly sampled unit from the blend, i.e. $\sigma_{total} = \sigma_{sample}$. The standard deviation corresponding to replicates is fixed, $\sigma_{rep} = 1$.

tance. The reason is that the PQRI criteria allows for sampling bias of content. Allowing the mean content of samples to deviate from LC, reduces the risk of rejecting homogeneous batches with a mean content close to LC in cases of sampling bias. This is in contrast to the USP criteria for tablets which besides uniformity also controls deviation from LC (see Appendix D and Appendix C).

Further it is seen that the level curves are rather smooth and almost linear and almost independent of the mean content. In these simulations the level curves depend weakly on the mean content through $CV = \frac{\sigma_{total}}{\mu} = \frac{\sigma_{area}^2 + 1}{\mu}$. Thus the test is essentially a test on CV even though it also includes limits for individual observations.

In this situation with fixed variation between replicates and no variation between layers, there exists an unambiguous relation between the probability of passing the PQRI test and the total standard deviation in the batch. This is also seen from Figure 3 which is a sectional plane of the OC-surface in Figure 2 for $\mu = 100$. From superimposing Figure 3 on Figure 1 it is also seen that the total standard deviation in the batch obtained from combining the variation among areas and the replicate variation in accordance with expression (2) is identical to the the total standard deviation among 12 uncorrelated, identically (normally) distributed samples in Figure 1. This is not surprising as the areas from which each of the 12 samples in model 1 is sampled are uncorrelated.

As the OC-surface is only moderately dependent of the mean content in the batch as seen in Figure 2, the rest of the analysis in this appendix will for simplicity focus on blends with a mean content equal to LC.

3 Batches with large scale variation (variation between layers)

In this section batches with large scale variation, σ_{layer}^2 , (and fixed small scale variation, $\sigma_{rep} = 1$), but no medium scale variation, $\sigma_{area,hi}^2$, are considered.

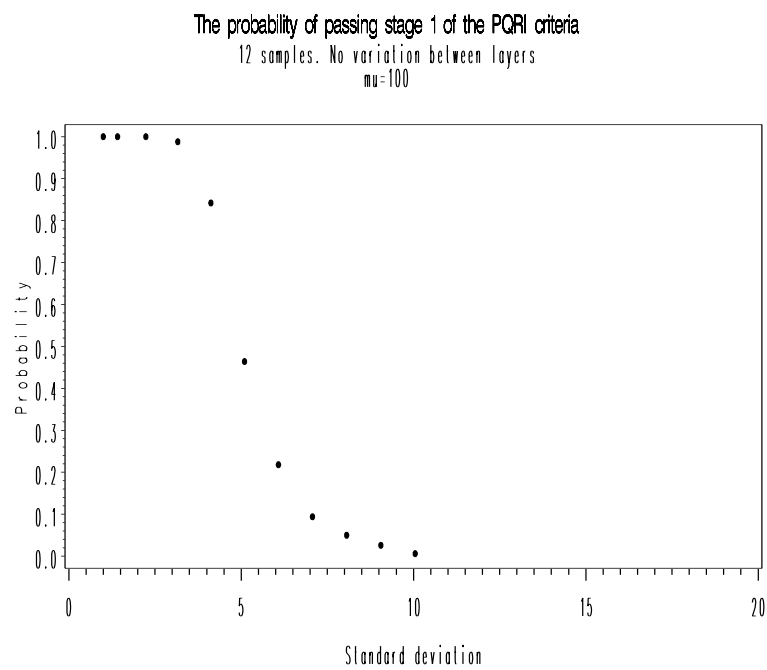


Fig. 3: The probability of passing stage 1 of the PQRI criteria in a situation with no variation between layers and fixed variation between replicates. The probability is plotted against the total standard deviation in the batch determined from expression (2) and correspond to a sampling plan with one sample from each of 12 different areas.

3.1 The total variation from the ANOVA table

With three layers and four samples from each layer a model for the content in a sample taking into account only large scale variation is

$$Y_{ij} = \mu + L_i + E_{ij}, \quad \begin{cases} i = 1, 2, 3 \\ j = 1, 2, 3, 4 \end{cases} . \quad (3)$$

Y_{ij} is the content of a sample, μ the mean content in the batch, L_i the effect corresponding to the layer from which the sample is taken, and E_{ij} is the error term also including the variation due to sampling. Further $L_i \in N(0, \sigma_{layer}^2)$ and $E_{ij} \in N(0, \sigma_{rep}^2)$.

The corresponding ANOVA table is

ANOVA Table			
Source of variation	Sums of squares	Degrees of freedom (df)	Expected mean square (EMS)
Between layers	SS_{layer}	3 - 1	$4\sigma_{layer}^2 + \sigma_{rep}^2$
Error	SS_{rep}	3(4-1)	σ_{rep}^2
Total	SS_{total}	12 - 1	

Table 3: The analysis of variance table corresponding to the model (3).

From Table 3 the estimate of the total variation in the batch is

$$\begin{aligned} \sigma_{total}^2 &= \frac{SS_{layer} + SS_{rep}}{df_{total}} \\ &= \frac{df_{layer} \times MSE_{layer} + df_{rep} \times MSE_{rep}}{df_{total}} \\ &= \frac{8\sigma_{layer}^2 + 2\sigma_{rep}^2 + 9\sigma_{rep}^2}{11} \\ &= \frac{8}{11}\sigma_{layer}^2 + \sigma_{rep}^2. \quad (4) \end{aligned}$$

The OC-surface is plotted as a function of the standard deviation corresponding to this expression with dots in Figure 4. μ is fixed at LC.

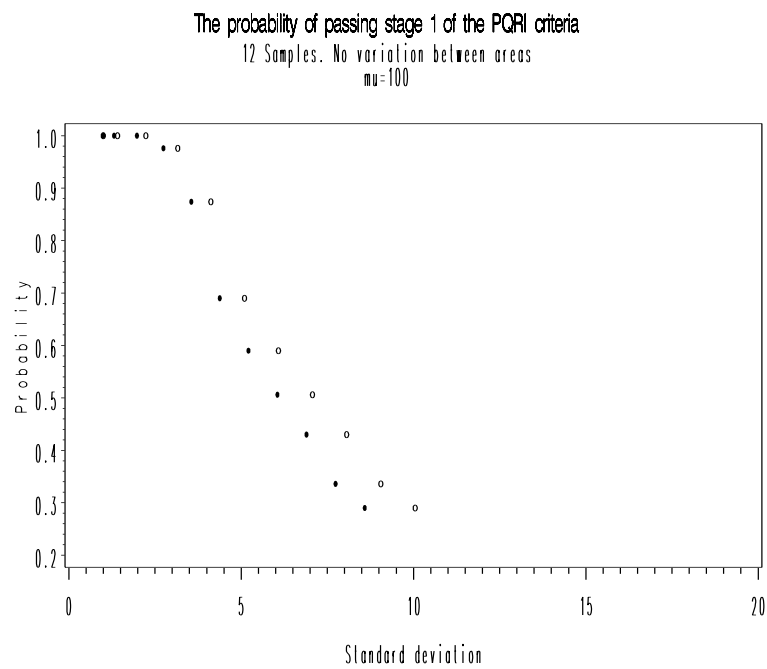


Fig. 4: The OC-surface plotted as a function of various measures of the total standard deviation in the batch for $\mu = LC$ and $\sigma_{rep} = 1$. The dotted line corresponds to the total standard deviation as derived from expression (4), $\sigma_{total} = \sqrt{\frac{8}{11}\sigma_{layer}^2 + \sigma_{rep}^2}$, and the circles correspond to the total standard deviation defined as the total standard deviation on a random unit sampled from the blend, $\sigma_{sample} = \sqrt{\sigma_{layer}^2 + \sigma_{rep}^2}$.

3.2 The variance on a sample from the batch

According to model (3) the variation on a sample from the blend is $\sigma_{sample}^2 = \sigma_{layer}^2 + \sigma_{rep}^2$. Thus, with the given sampling plan and under model (3) the expression for the total variation in the batch according to Table 3, σ_{total}^2 , and the expression for the variation on a randomly sampled unit from the blend, σ_{sample}^2 , are not identical. The variation on a sample from the blend is always the sum of the variance components corresponding to the given model, whereas the expression for the total variation derived from an ANOVA table is only a sum with all coefficients equal to one in a few cases, depending on the corresponding sampling plan.

As the dotted line in Figure 4 is more steep than the circled line it appears that in a batch with large scale but no medium scale variation the acceptance criteria is more efficient at separating batches with good quality from batches with a low quality when the measure of total variation is as derived from the ANOVA table rather than from the expression for the variance on a randomly sampled unit from the batch.

Further as both lines in Figure 4 is less steep than the line in Figure 3 it is seen that under a sampling plan with 12 observations from three layers and four areas within each layer the acceptance criteria is more sensitive in the case of medium scale variation but no large scale variation than in the case of large scale variation and no medium scale variation in the blend.

4 Batches with both large and medium scale variation

In real batches both large and medium scale variation may occur at the same time. Therefore, in this section expressions for the total blend homogeneity is derived for batches with both large and medium scale variation (and fixed small scale variation).

4.1 The total variation from the ANOVA table

Under a sampling plan with three layers, four areas within each layer and (in accordance with the first stage of the PQRI procedure) only one sample from each area, the hierarchical model (2) from Appendix A reduces to

$$Y_{ijk} = \mu + L_i + A(L)_{j(i)} + E_{ij}, \quad \begin{cases} i = 1, 2, 3 \\ j = 1, 2, 3, 4 \end{cases} \quad (5)$$

L_i is the effect from layer i , $A(L)_{j(i)}$ is the effect from the j 'th area within layer i and E_{ij} is the error term also including the variation due to sampling. Further it is assumed that $L_i \in N(0, \sigma_{layer}^2)$, $A(L)_{j(i)} \in N(0, \sigma_{area,hi}^2)$ and $E_{ij} \in N(0, \sigma_{rep}^2)$.

The corresponding ANOVA table is

ANOVA Table			
Source of variation	Sums of squares	Degrees of freedom (df)	Expected mean square (EMS)
Between layers	SS_{layer}	3 - 1	$4\sigma_{layer}^2 + \sigma_{area,hi}^2 + \sigma_{rep}^2$
Between areas, within a layer	$SS_{area,hi}$	3(4 - 1)	$\sigma_{area,hi}^2 + \sigma_{rep}^2$
Error	SS_{rep}	0	σ_{rep}^2
Total	SS_{total}	12 - 1	

Table 4: The analysis of variance table corresponding to the hierarchical model (5).

In analogy with the derivations of σ_{total}^2 in the previous sections, the expression for σ_{total}^2 derived from Table 4 is $\sigma_{total}^2 = \frac{8}{11}\sigma_{layer}^2 + \sigma_{area,hi}^2 + \sigma_{rep}^2$. The OC-surface is plotted as a function of the corresponding standard deviation, σ_{total} , with dots in Figure 5 to Figure 7. The difference between the three figures is that the mean content is fixed at respectively 90%, 100% and 110 % of LC. It is seen that the plots in the three figures are very alike in agreement with the fact that the acceptance criteria is only moderately dependent on the value of the mean content in the batch, μ .

It is also seen that for this OC-surface there is not an unambiguous relation

between the total variation in the blend and the probability of passing the test². Figure 5 to Figure 7 is the resulting plot from superimposing all plots for specified values of respectively variation between areas and variation between layers. If σ_{area} is fixed at same value the plot is a parallel displaced version of Figure 4. If σ_{layer} is fixed at same value the plot is a parallel displaced version of Figure 3.

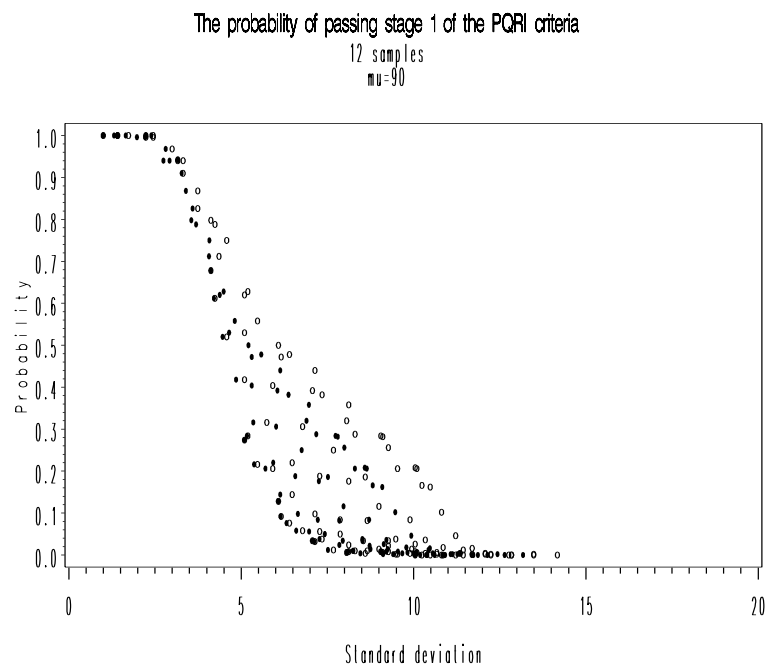


Fig. 5: The dots correspond to the total standard deviation as derived from $\sigma_{total}^2 = \frac{8}{11}\sigma_{layer}^2 + \sigma_{area,hi}^2 + \sigma_{rep}^2$, and the circles correspond to the total standard deviation on a randomly sampled unit from the blend, $\sigma_{sample}^2 = \sigma_{layer}^2 + \sigma_{area,hi}^2 + \sigma_{rep}^2$. The mean content in the batch is 90% of LC.

It should be remarked that if there had been three replicates per area as in the batches simulated in Section 2 in Appendix A, the expression for the total variation in the batch derived from the corresponding ANOVA table would be $\sigma_{total}^2 = \frac{8}{35}\sigma_{layer}^2 + \frac{33}{35}\sigma_{area,hi}^2 + \sigma_{rep}^2$.

²In the simulations $\sigma_{area,hi}$ and σ_{layer} vary between 0 and 10. Had larger values for the two parameters been included in the simulations the OC-band would have been even broader in these figures

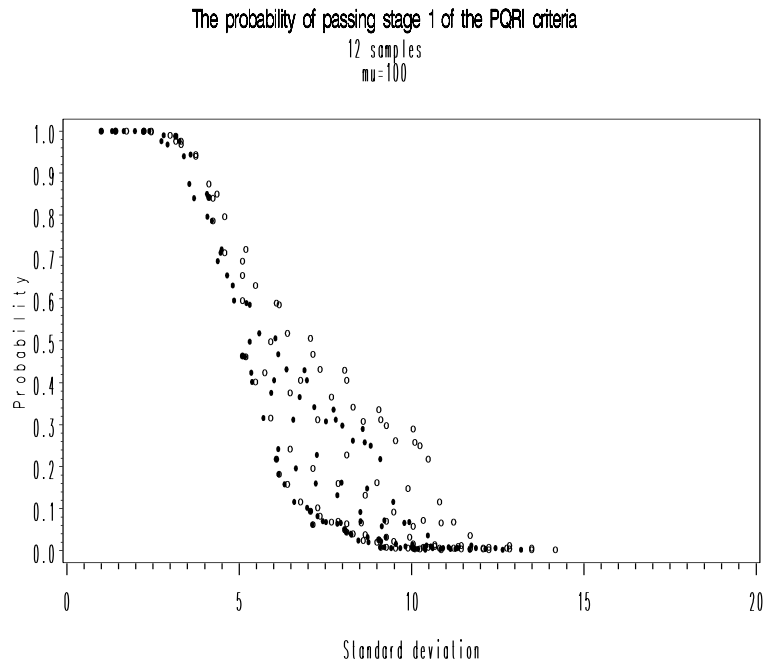


Fig. 6: The dots correspond to the total standard deviation as derived from $\sigma_{total}^2 = \frac{8}{11}\sigma_{layer}^2 + \sigma_{area,hi}^2 + \sigma_{rep}^2$, and the circles correspond to the total standard deviation on a randomly sampled unit from the blend, $\sigma_{sample}^2 = \sigma_{layer}^2 + \sigma_{area,hi}^2 + \sigma_{rep}^2$. The mean content in the batch is 100% of LC.

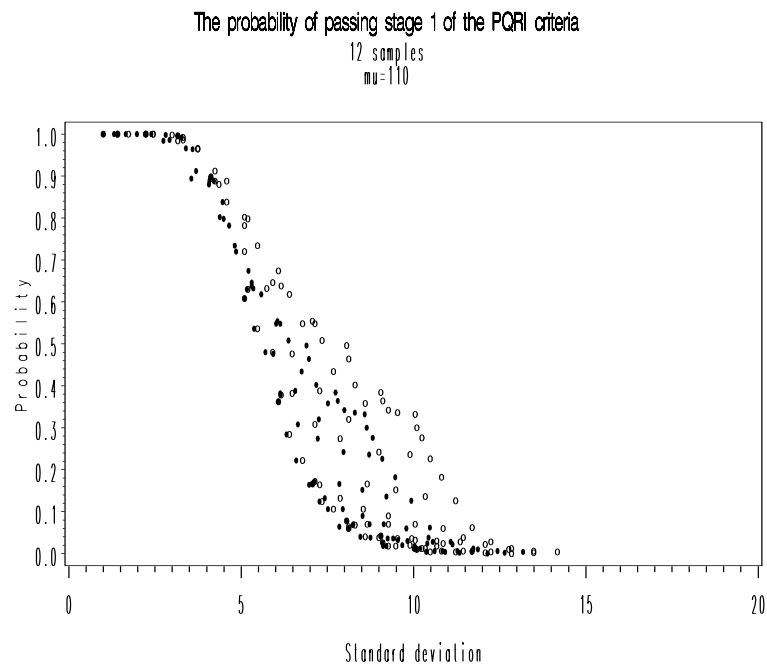


Fig. 7: The dots correspond to the total standard deviation as derived from $\sigma_{total}^2 = \frac{8}{11}\sigma_{layer}^2 + \sigma_{area,hi}^2 + \sigma_{rep}^2$, and the circles correspond to the total standard deviation on a randomly sampled unit from the blend, $\sigma_{sample}^2 = \sigma_{layer}^2 + \sigma_{area,hi}^2 + \sigma_{rep}^2$. The mean content in the batch is 110% of LC.

The OC-surface is plotted as a function of the standard deviation corresponding to this expression with dots in Figure 8. Comparing the dotted OC-surfaces in Figure 6 and Figure 8 shows that under this hierarchical sampling plan the relation between the total variation in the blend and the probability for a batch to pass the acceptance criteria is even less clear with three replicates per area than under the hierarchical sampling plan with only one observation per area.

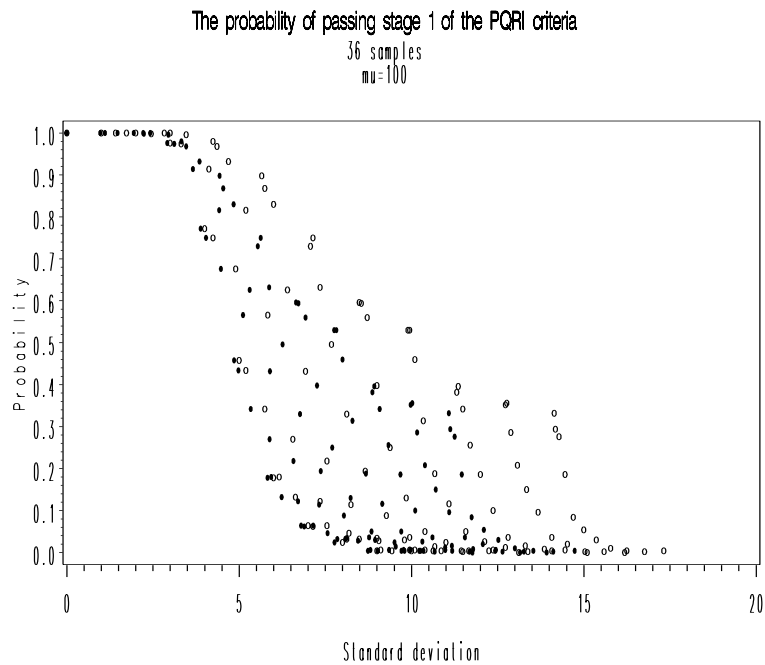


Fig. 8: The OC-surface for a hierarchical sampling plan with a total of 36 samples. The dots correspond to the total standard deviation derived from $\sigma_{total} = \sqrt{\frac{8}{35}\sigma_{layer}^2 + \frac{33}{35}\sigma_{area,hi}^2 + \sigma_{rep}^2}$, and the circles correspond to the total standard deviation on a randomly sampled unit from the blend, $\sigma_{sample} = \sqrt{\sigma_{layer}^2 + \sigma_{area,hi}^2 + \sigma_{rep}^2}$. The mean content in the batch is 100% of LC.

Figure 9 is similar to Figure 8 except that the criteria for the individual tablets are disregarded. It is seen that as long as the criteria for the sample coefficient of variation is included, the criteria on the individual tablets does not have a large effect on which batches are accepted by the acceptance criteria.

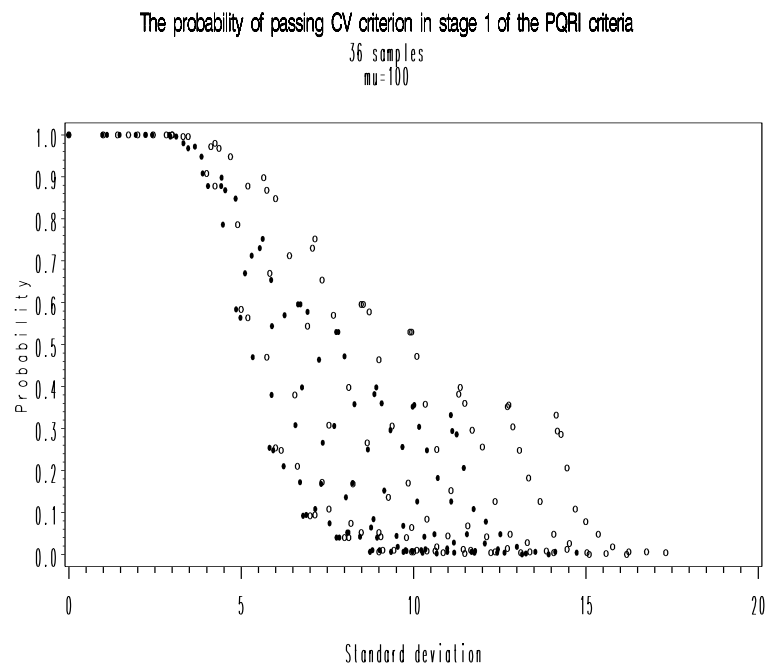


Fig. 9: The figure is similar to Figure 8 except that the criteria for the individual tablets are left out. The dots correspond to σ_{total} and the circles correspond to σ_{sample} .

4.2 The variance on a sample from the batch

The total variance of a randomly sampled unit from the hierarchical model (5) is $\sigma_{sample}^2 = \sigma_{layer}^2 + \sigma_{area,hi}^2 + \sigma_{rep}^2$.

The circles in Figure 5 to Figure 9 corresponds to the total variation in the batch expressed as σ_{sample} . As the OC-bands corresponding to dots in general is more narrow than the OC-bands corresponding to circles it is seen that both under the hierarchical model with 12 samples and the hierarchical model with 36 samples, the acceptance criterion is more unambiguous related to the measure of total variation, σ_{total} , than to σ_{sample} .

4.3 The direct relation between large/medium scale variation and the acceptance criteria

In the previous sections various expressions for the total variation in the blend has been derived as a function of the variation between respectively layers, areas and between replicates. These expressions are not always straight forward and they are dependent on the specific sampling plane. Thus in case of both medium and large scale variation (and fixed variation between replicates) in the blend it may be more straight forward to investigate the OC-surface for the acceptance criteria directly as a function of σ_{layer} and $\sigma_{area,hi}$ respectively. In Figure 10 through Figure 12 this relation is plotted for the hierarchical model with 12 samples. In the three figures μ is fixed at respectively 85 % LC, 100 % LC and 115 % LC. The variation between replicates is fixed, $\sigma_{rep} = 1$.

5 Discussion and conclusion

The aim of this appendix is to investigate how a concept of the total variation in the blend, can be interpreted in case of inhomogeneities in the blend.

The interpretation of the total variation is relevant as acceptance criteria for blend uniformity is based on considerations on the total variation.

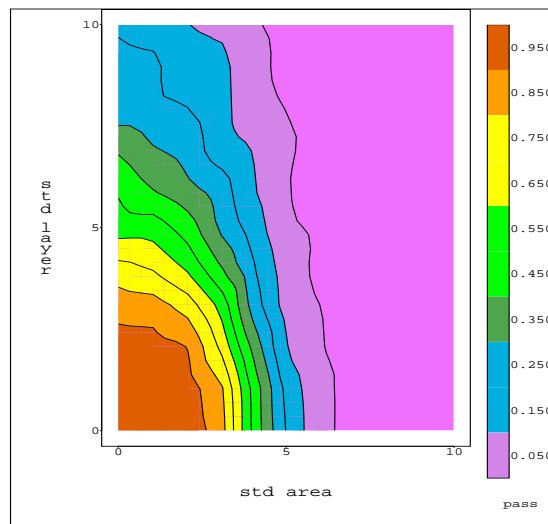


Fig. 10: Level curves for the OC-function of the first stage of the acceptance criteria for blend uniformity suggested by PQRI. The mean content in the batch, μ , is 85 % LC and the variation between replicates is fixed, $\sigma_{rep} = 1$.

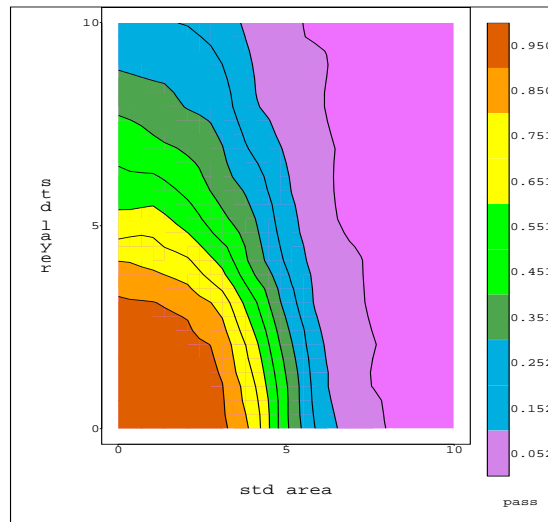


Fig. 11: Level curves for the OC-function of the first stage of the acceptance criteria for blend uniformity suggested by PQRI. The mean content in the batch, μ , is 100 % LC and the variation between replicates is fixed, $\sigma_{rep} = 1$.

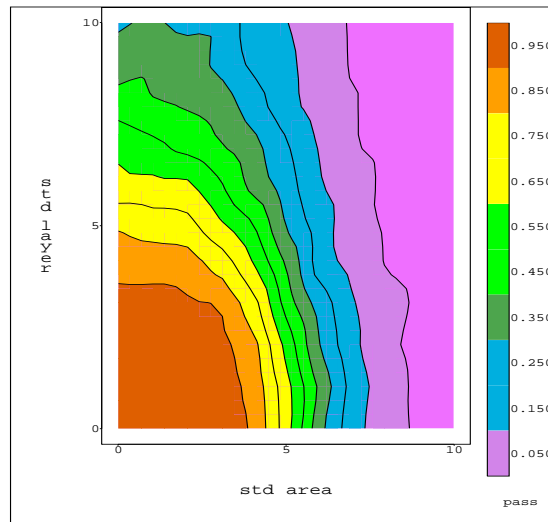


Fig. 12: Level curves for the OC-function of the first stage of the acceptance criteria for blend uniformity suggested by PQRI. The mean content in the batch, μ , is 115 % LC and the variation between replicates is fixed, $\sigma_{rep} = 1$.

In case of no variation between layers and under a hierarchical sampling plan with only one replicate per area, the measure for the total variation in the blend derived from the analysis of variance table and the total variation defined as the variance on a random sample from the blend are identical. The reason is that in this case the samples are uncorrelated. However, in presence of large scale variation or when collecting more than one replicate from each area in presence of medium scale variation the two expressions are not identical.

In general the total variation is defined from the analysis of variance table leads to more effective (steeper OC-surface) and less ambiguous (more narrow OC-band) acceptance criteria than the total variation defined as the total variation on a random sample from the batch. However, the expression for total variation defined as the total variance from the analysis of variance table, depends on the sampling plan used.

Paper C

On a test for content uniformity in pharmaceutical products

Presented at the First Annual ENBIS Conference, Oslo 2001

C

Abstract

A proposed amendment of the current procedure in the US Pharmacopeia for testing content uniformity in tablets is a two-stage sampling plan to be applied to each batch during validation.

In statistical terms, the procedure at each stage is a combination of a test by attributes with acceptance number zero with “specification limits” $\pm 25\%$ around the target value, and a test by variables with “specification limits” $\pm 16.5\%$ around the target value. In the paper we discuss the statistical properties of this procedure.

Keywords

Specification limits, inspection by attributes, inspection by variables, OC-curve, acceptance sampling.

1 Introduction

To assure the therapeutic utility of dosage units such as tablets, the drug content of each unit in a batch should not deviate too much from a chosen target value, e.g. label claim (LC).

A procedure to control the variability (uniformity) of the content of compressed tablets in a batch is the test for content uniformity specified by the US authorities [26].

The current test for content uniformity is a two-stage test including tests by attributes for the content of single tablets in the sample expressed relatively to the label claim (LC) as well as a limit for the sample coefficient of variation.

Pharmaceutical companies often sell their products in several regions of the world and therefore have to comply with the requirements in each of the countries and areas in which they sell their products. In order to reduce the need to

duplicate the testing carried out during the research and development of new drugs, efforts are made to achieve greater harmonisation in the interpretation and application of technical guidelines and requirements for product registration.

Regarding testing for content uniformity this has so far resulted in a proposal from the US authorities for an alternative to the current test procedure.

The proposed test procedure, which is the subject of this article, is a two-stage test using in each stage a test by attributes on the content of the individual tablets combined with a test by variables.

2 The proposed test

2.1 Historical notes

The currently valid requirement from the US authorities concerning content uniformity is termed USP 24 [26]. This test is a two stage sampling plan with ten tablets tested on stage 1 and further twenty tablets tested on stage 2. On each stage the test has two requirements to the drug content of individual tablets in the sample: At most one tablet is allowed to be outside the limits 1.0 ± 0.15 LC and no tablet outside the limits 1.0 ± 0.25 LC. Actually, the limits for the individual tablets depend on the sample mean. Further the sample coefficient of variation shall be less than 6 % and 7.8 % respectively for stage 1 and stage 2.

The first proposed amendment to the requirements in the form described above was moved in 1997 [27]. The amendment has been commented and discussed over the last years for the present resulting in a proposed test described in Pharmacopeial Forum [28]. The version of the proposed test under concern in this article is described in Pharmacopeial Forum [29]. However, the content of the various versions of the proposed test are essentially the same, and somewhat different from the the approach in the current requirement.

The current USP specification controls the content variability by establishing a limit on the sample coefficient of variation. Further the test includes limits

for the content in individual tablets in the sample. In this procedure the mean content is not directly controlled. This is in line with the test being a test for uniformity, i.e. variability, and not explicitly for mean content. The proposed test, however, controls content variability by means of a linear combination of the sample standard deviation and the absolute mean deviation from the target. Also in this procedure a test for the content in individual tablets is included. With the linear combination including sample standard deviation and mean in the new procedure not only the variability but also the mean is controlled. Thus, what historically started as a test for uniformity, still is termed a test for content uniformity, but now also controls the mean content.

The details of the proposed test, see [29], are described in the following.

2.2 Description of the proposed test

The proposed test is a two stage test which combines a criterion on the sample mean and standard deviation, (\bar{x}, s) , with a criterion on individual tablets. The details of the testing procedure is outlined schematically in Figure 1¹.

Before the testing procedure is started 30 tablets are selected from the batch.

On stage 1 of the testing procedure a sample of 10 tablets is investigated. Test stage 1 may either lead to acceptance, rejection, or to further testing on stage 2. The batch of tablets is accepted at this stage if the sample mean and standard deviation, (\bar{x}, s) , are within the acceptance limits, and all individual tablets are within limits for individual tablets specified by the procedure. These limits depend on the sample mean. The batch is rejected if the content of one or more individual tablet in the sample is outside the limits for individual tablets. If no individual tablets is outside the limits but the combination of sample mean and sample standard deviation is outside the acceptance limits, further testing is performed in accordance with the criteria for test stage 2.

The sample used for test stage 2 consists of the 10 tablets from stage 1 combined with the remaining 20 tablets. Stage 2 results in either acceptance or

¹For simplicity only the part of the proposed test that concerns compressed tablets with a rubric mean less than 101.5 percent of label claim is considered. The rubric mean is the average of the shelf limits specified in the potency definition of the drug.

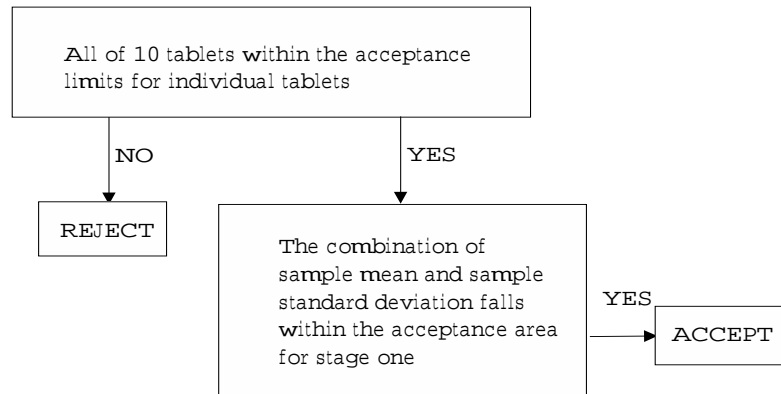
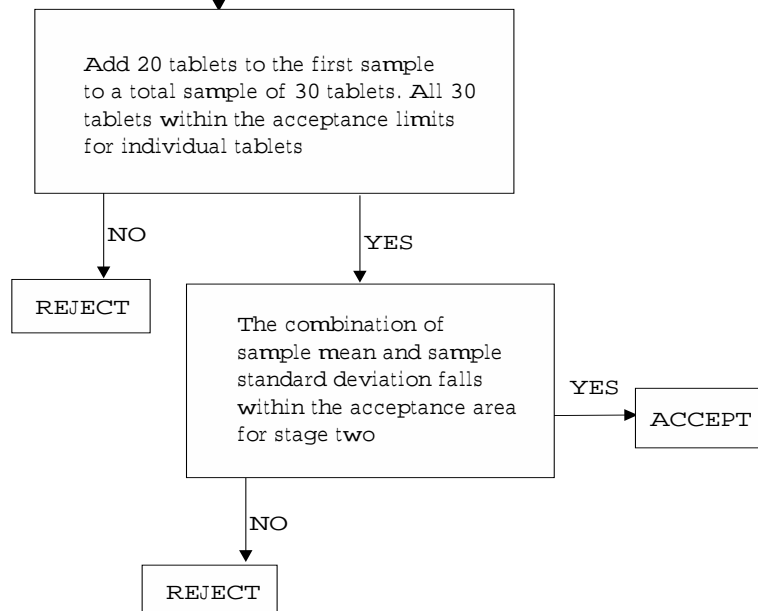
Stage 1**Stage 2**

Fig. 1: Schematic representation of the proposed test for content uniformity.

\bar{x}	Acceptance limits for s	Acceptance limits for individual tablets in the sample
Stage 1		
$\bar{x} \leq 0.835$	-	0.74 – 1.23
$0.835 \leq \bar{x} \leq 0.985$	$s \leq (\bar{x} - 0.835)/2.3$	0.74 – 1.23
$0.985 \leq \bar{x} \leq 1.015$	$s \leq 0.065$	$0.75\bar{x} - 1.25\bar{x}$
$1.015 \leq \bar{x} \leq 1.165$	$s \leq (1.165 - \bar{x})/2.3$	0.76 – 1.27
$1.165 \leq \bar{x}$	-	0.76 – 1.27
Stage 2		
$\bar{x} \leq 0.835$	-	0.74 – 1.23
$0.835 \leq \bar{x} \leq 0.985$	$s \leq (\bar{x} - 0.835)/2.0$	0.74 – 1.23
$0.985 \leq \bar{x} \leq 1.015$	$s \leq 0.075$	$0.75\bar{x} - 1.25\bar{x}$
$1.015 \leq \bar{x} \leq 1.165$	$s \leq (1.165 - \bar{x})/2.0$	0.76 – 1.27
$1.165 \leq \bar{x}$	-	0.76 – 1.27

Table 1: Acceptance limits, measured relatively to label claim (LC), for the sample standard deviation and for individual tablets. In general the limits are functions of the sample mean. If $\bar{x} \leq 0.835$ LC or $\bar{x} \geq 1.165$ LC the batch is rejected. \bar{x} and s are based on a sample of 10 tablets on stage 1 and the total sample of 30 tablets on stage 2.

rejection of the batch. The batch is accepted if criteria for individual tablets as well as criteria for the sample mean and sample standard deviation are satisfied. Otherwise the batch is rejected.

Acceptance limits for individual tablets and the sample standard deviation as a function of the sample mean are shown in Table 1. The corresponding acceptance area in the (\bar{x}, s) -plane for the test on sample mean and standard deviation is shown in Figure 2.

3 Properties of the proposed test

The proposed test evaluates a batch on the basis of the content and the variation of the active component (drug) in the batch. Therefore, assuming a normal distribution of tablet content, the properties of the test is described by the

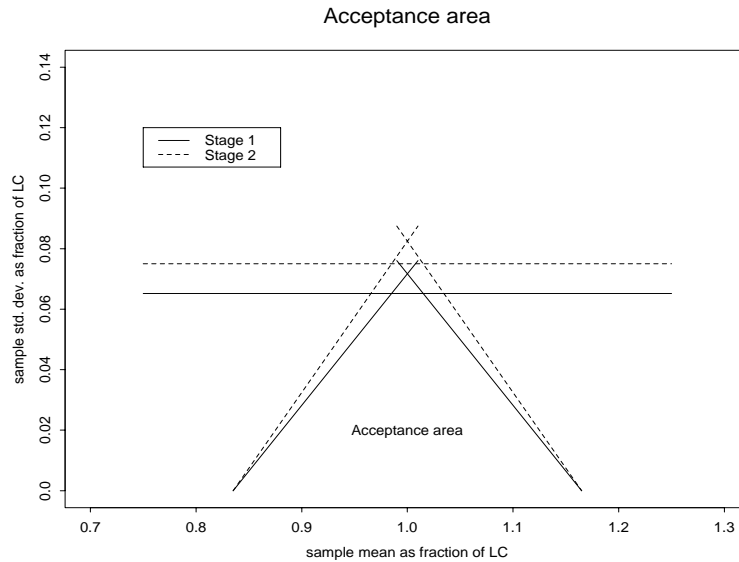


Fig. 2: Acceptance area for sample mean content and standard deviation.

probability for a batch to be accepted by the test as a function of the mean and standard deviation, (μ, σ) , of the content in the batch. Figure 3 shows the level-curves for the OC-surface for the test. The combination of requirements on individual tablets with requirements on \bar{x} and s complicates the derivation of an exact analytical expression for the OC-surface. Instead the OC-surface is based on Monte Carlo simulations in S-Plus version 6 [30]. 3000 simulations were performed for each of various combination of μ and σ and level curves have been determined by interpolation.

In this article the influence of measurement errors on the OC-surface is not taken into account. For a discussion of the effect of measurement errors in acceptance sampling see [31] and [32].

3.1 Description of the OC-surface of the test

From Figure 3 it is seen that when the standard deviation, σ , exceeds 0.095 LC the probability of accepting the batch is less than 0.1 no matter what the mean content, μ , is. Similarly, when the mean content is less than 0.835 LC or larger than 1.165 LC the probability of accepting the batch is less than 0.1, no matter

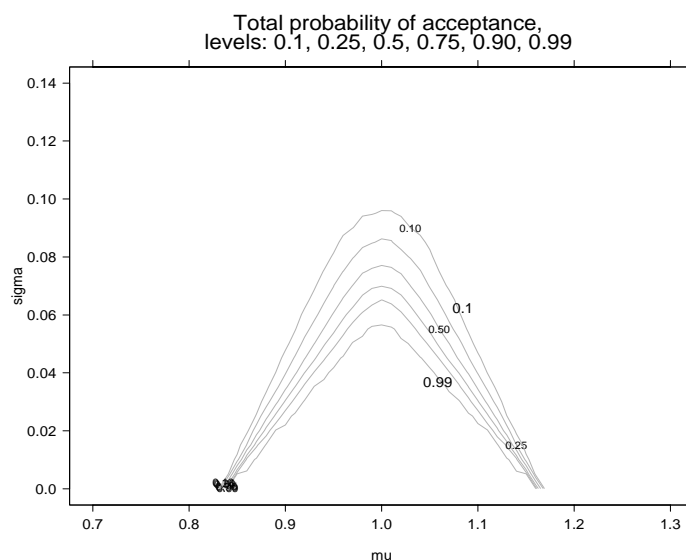


Fig. 3: Level-curves for the OC-surface for the test. The curves are based on 3000 simulations for each lattice point (μ, σ) .

how small the standard deviation in the batch is.

Within the innermost triangular-like shape in Figure 3 the probability of accepting the batch is 0.99 or larger.

3.2 “Specification limits” for individual tablets

As a single quality measure for a batch, one might consider the proportion of tablets in the batch with a content outside some specified range of values. The U.S authorities have not explicitly specified limiting values for the content of tablets in the batch, nor is an unacceptable proportion specified. However, the acceptance limits for individual tablets in Table 1 suggests that it is not considered satisfactory for the content in a tablet to deviate more than 25% from label claim. Figure 4 shows the proportion of tablets with a content outside 1 ± 0.25 LC in a batch as a function of μ and σ .

Comparing the proportion of tablets in a batch outside these limits with the probability for acceptance of the batch is a way to evaluate the properties of

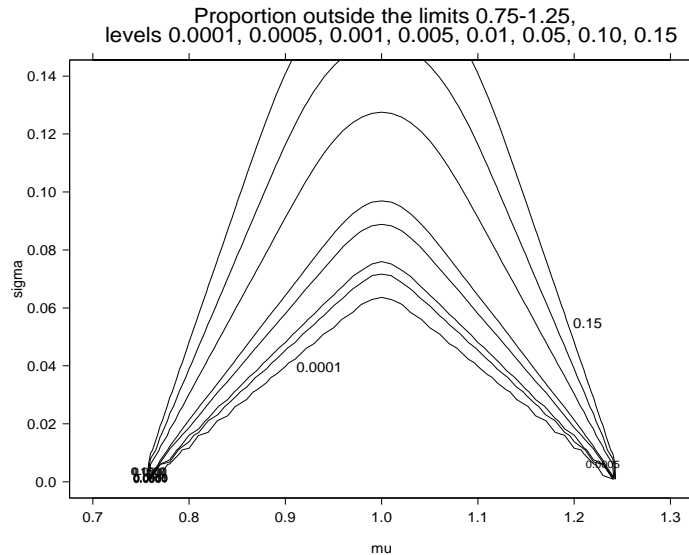


Fig. 4: Level-curves for the proportion of tablets in a batch outside the limits 1 ± 0.25 LC.

the test: Batches with a large proportion of tablets outside these limits should have a high probability of being rejected by the test, and batches with a very small proportion of tablets outside the limits should have a high probability of acceptance.

Superimposing Figure 3, the probability for a batch to be accepted by the test, upon Figure 4, the quality of the batch, reveals that batches with a probability of 0.9 of being accepted have a proportion of tablets outside the specification of less than 0.0001.

As the curves in the two figures do not have the same shape it is not possible to determine the probability of rejecting a batch as a single-valued function of the proportion of tablets outside the specification, i.e. it is not possible unambiguously to determine which proportion of unacceptable tablets in the batch will lead to rejection with a given probability. As an example in some cases batches with a proportion of 0.0001 outside the specification is rejected with a probability of more than 0.9. However, in other cases, depending on the value of μ , such batches have a probability of 0.99 of acceptance.

The acceptance limits for the sample standard deviation as a function of the

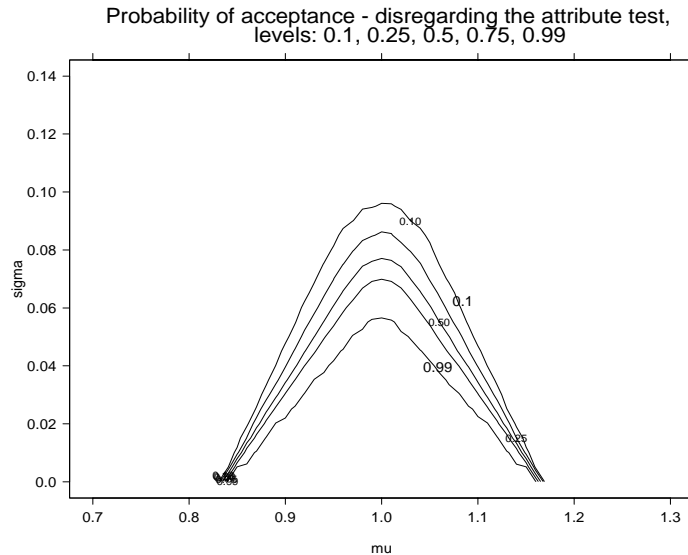


Fig. 6: Level-curves for the OC-surface for the test - disregarding the attribute test. The level-curves are based on 3000 simulations for each lattice point (μ, σ) .

However, the discriminatory ability is the result of the combined properties of the individual elements of the test. These elements are a test by attributes and a test by variables in each of two stages of the test. The test by attributes is (essentially) a test for the proportion of tablets outside 1 ± 0.25 LC. The shape of the acceptance region for (\bar{x}, s) in Figure 2, resembles the acceptance region for the test by variables controlling the proportion of tablets outside a given specification as derived by Lieberman and Resnikoff [33] and described by Schilling [12]. This suggests that the proposed test by variables for practical purposes controls the proportion of tablets outside 1 ± 0.165 LC. As the specification 1 ± 0.25 LC is less restrictive than the specification 1 ± 0.165 LC it is indicated that the test by variables is the effective part of the test.

To further investigate the effect of respectively the test by variables and the test by attributes the level-curves for the OC-surface is plotted for a test procedure disregarding the attribute test. These level-curves are shown in Figure 6. Superimposing this figure upon Figure 3, the level-curves for the total test, reveals that the probability for a batch to be accepted by the total test virtually does not depend on whether the attribute test is included in the test or not. This is in line with the fact that under the assumption of normality (\bar{x}, s) are jointly sufficient for (μ, σ) , and hence the conditional distribution of the content of individual

tablets in the sample for a given combination (\bar{x}, σ) does not depend on the value of (μ, σ) .

Now, consider the two stages of the test. Figure 7 shows the probability of invoking test stage 2. Batches corresponding to a (μ, σ) -combination in the inner triangular area have a high probability of acceptance on stage 1. Batches corresponding to a (μ, σ) -combination outside the plotted level-curves have a high probability of being rejected on stage 1, whereas batches corresponding to a (μ, σ) -combination in the 'sausage'-shape in the middle have a high probability of invoking test stage 2. For these latter batches 30 tablets instead of 10 shall be analysed, i.e. such batches are expensive regarding time and other resources to the testing and chemical analysis.

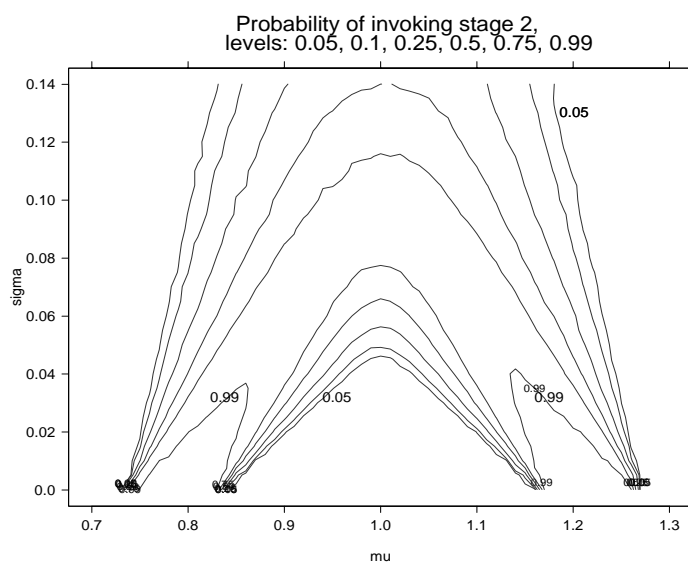


Fig. 7: The probability of invoking stage 2 when both the variable and the attribute tests are included in the test procedure. The level-curves are based on 3000 simulations for each lattice point (μ, σ) .

Figure 8 shows the probability of invoking test stage 2 when the attribute test on individual tablets is disregarded in the procedure. Batches corresponding to a (μ, σ) -combination in the inner triangular area have a high probability of acceptance on test stage 1. However, disregarding the test for individual tablets rejection on stage 1 is not possible, implying that all batches corresponding to a (μ, σ) -combination outside the inner triangular area would have a high probability of invoking test stage 2, hence requiring the efforts of further testing.

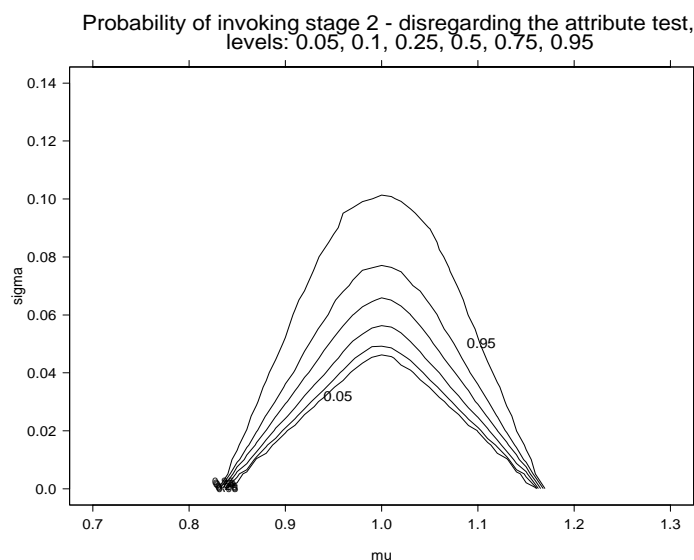


Fig. 8: The probability of invoking stage 2 - disregarding the attribute test. The level-curves are based on 3000 simulations for each lattice point (μ, σ) .

Thus, a very important effect of the test by attributes is to reduce resources to the testing and the chemical analysis by rejecting most of the unacceptable batches already on stage 1. Actually, it is possible to design criteria for rejection on stage 1 that are even more effective than the above attribute test. Schilling [12] describes a double sampling plan by variables that allows for rejection on stage 1.

Other benefits of including the test by attributes in the test procedure is that the test provides robustness of the procedure in case of a non-normal distribution of the tablet content. The discriminating properties of the test in the case of non-normal distribution of the tablet content have not been investigated. The limits for individual tablets may in these situations serve as a safety precaution.

Finally the test by attributes may serve a psychological purpose. The reason is that the use of a test by variables - without a test by attributes - might (in extreme cases) lead to acceptance of a batch even though tablets of a very extreme quality is found in the sample. The attribute specification in the proposed test prevents such a situation.

4 Comparison to other test procedures

It is well understood in most commercial sectors that complete testing of product is resource-demanding, and - in case of destructive testing - even impossible, and therefore there is a long industrial tradition for acceptance of product based upon inspection of a sample. Statistical theories for acceptance sampling by attributes dates back to the pioneering work by Molina, Dodge and Romig at Bell Telephone Laboratories in the 1920's and the theoretical basis for acceptance sampling by variables was given by Lieberman and Resnikoff [33] in the 1950's. The current international standards for acceptance sampling by attributes, ISO 2859 [34], has been widely accepted by industry as pragmatic rules to be used in agreements by two parties for releasing product after inspection of only a limited sample of the product. Because of the greater efficiency of sampling by variables, the complementary standard, ISO 3951 [35], for acceptance sampling by variables is also used in many industrial relations.

Although the current trend in quality management is to shift the focus from the final product inspection to the monitoring of the process, acceptance sampling may still serve a purpose as part of quality control procedures as described e.g. in the quality management standard QS 9000 [36] developed by the automotive industry and the US military standard, MIL-STD-1916 [37] for acceptance of product. However, in line with the quest for "zero defects", the acceptance sampling plans suggested by these standards are plans with acceptance number zero ("accept zero plans") for sampling by attributes, and equivalent plans for sampling by variables. Although the assumption of normally distributed item characteristics is seldomly questioned in the calculation of process capability and performance measures, there is however, some reluctance towards using variables sampling plans, partly because of the underlying assumption of a normal distribution, and partly because the use of a variables sampling plan might (in extreme cases) lead to acceptance of product even when a nonconforming item is found in the sample. Such situations are in apparent conflict with the "accept zero" philosophy.

Although acceptance sampling of product in commercial transactions between two parties has a more pragmatic purpose than sampling inspection for regulatory purposes they share the common goal of establishing evidence that product is of a satisfactory quality.

Thus, the EC legislation on weights and measures of prepackaged goods [38]

and [39] lays down a sampling procedure to be used by the authorities for verifying compliance to the labelled content. The procedure is a combination of a test by attributes for the proportion of packages with content less than specified, and a separate test by variables leading to rejectance when the sample average is significantly smaller than the specified content in the packages. Although the test by variables is only concerned with the mean content, the combined result of the two tests is an economic incentive to the producer to maintain a small variance in order to avoid overfill.

However, in case of drugs, “overfill” is of just as much concern as “underfill”, and therefore positive as well as negative deviations from label claim are undesirable.

Standard acceptance sampling plans for industrial use are based upon the usual industrial practice of setting up specifications for the final product (and all previous stages). Such specification limits, or tolerances, facilitate the communication between the parties and provide a basis for assessing quality simply by verifying compliance to specifications.

In the USP-proposal no explicit specification limits for the content of individual tablets has been laid down. Thus, in contrast to standard acceptance sampling procedures for product, the quality requirement to the product is specified only through the acceptance sampling procedure.

However, our analysis of the operations characteristics of the procedure provides some guidance to the producer and shows some analogies to standard procedures. Thus, the proposed procedure in essence is a procedure for controlling the proportion of tablets with content outside the 1 ± 0.165 LC, combined with an “accept zero” plan corresponding to the limits 1 ± 0.25 LC.

5 Conclusion

As a result of the efforts of achieving international harmonisation in the interpretation and application of technical guidelines and requirements for product registration the procedure for testing of content uniformity in tablets is under revision.

The proposed procedures all include both a test by attributes and a test by variables whereas the current procedure includes test by attributes and a test of the relative coefficient of variation.

In this article the discriminating properties of the proposed test procedure have been assessed. Under the assumption of normally distributed content in the tablets the qualities, (μ, σ) , leading to acceptance of a batch have been delineated. The analysis in this article reveals that the proposed procedure in essence is a procedure for controlling the proportion of tablets with content outside the 1 ± 0.165 LC, combined with an “accept zero” plan corresponding to the limits 1 ± 0.25 LC.

Further the effect of individual elements of the test procedure have been assessed.

The acceptance of a batch is determined by the test by variables. The function of the test by attributes is to reduce the resources to the testing and chemical analysis, to reject unsatisfactory batches in situations of non-normally distributed content in the tablets and finally it has a psychological effect as the use of a variables sampling plan might (in extreme cases) lead to acceptance of product even when a tablet of very extreme content is found in the sample.

Paper D

Statistical tests for uniformity of blend/dose samples

D

1 Introduction

Pharmaceutical companies are legally required to manufacture their products using current Good Manufacturing Practices (cGMPs) as defined, e.g. in documents from regulatory authorities. In pharmaceutical production as well as in other industrial sectors an important goal of good manufacturing practice is to control variability in the characteristics of the end product, and therefore good manufacturing practice implies monitoring of processes that may be responsible for causing variability in the characteristics of the end product, see e.g. the guidelines provided in the International Standard ISO 11462-1, [40].

In pharmaceutical production of tablets, a key process is the blending process producing the powder mix, and therefore most legal requirements prescribe control of blend processes with the purpose of demonstrating that a satisfactory degree of mixing has been achieved. Also, requirements on the monitoring of the final product (tablets) are provided with the purpose of demonstrating that the drug content of each unit in a lot is distributed within a narrow range around the label claim.

In the last decade there has been an increasing interest as well among pharmaceutical manufacturers as in regulatory agencies to clarify and standardize cGMP procedures for demonstrating blend uniformity, see the discussion in PDA Technical Report No. 25 [22] and the recent review by Berman [2], and to harmonize requirements on final product testing, see the series of proposed amendments to the United States Pharmacopeia tests for uniformity of dosage units, [27], [28].

As such control and monitoring procedures are based upon samples from the blend, or from the batch of tablets, there is an inherent uncertainty concerning the actual dispersion in the blend or batch being sampled. Therefore, the assurance provided by such procedures is of a statistical nature (i.e. depending on the pattern of variation in the entity under test), and in order to assess the influence of the uncertainty due to sampling the properties of the procedures may therefore be assessed using statistical concepts and techniques under due consideration to the potential sources of variation in the processes being mon-

itored.

In industrial and commercial practice, product requirements are often formulated as requirements on quantifiable characteristics of the product. Such requirements are most appropriately formulated as *specifications* for individual units of product, but may also include specifications for such batch or process characteristics as batch fraction nonconforming or standard deviation between units in the batch. The International Standard ISO 10576 [41] provides general guidelines on drafting specifications as well for commercial as for governmental regulatory purposes.

In particular, in ISO 10576 it is advised to separate the issue of designating specifications (i.e. range of permissible values) for a product, or a process from the issue of designating acceptance criteria to be used for assessment of conformity to the specifications. This facilitates the discussion of conformity assessment procedures in situations involving measurement- or sampling uncertainty, and allows for declarations of conformity (or nonconformity) that do not depend on the particular choice of measurement- or sampling method.

Thus, to be in line with this recommendation, requirements to e.g. uniformity of the blend and of doses should be formulated in terms of blend and batch characteristics, like mean dose content and standard deviation between dosage units in the batch, proportion of dosage units in the batch exceeding specified limits, etc. Although such a distinction between product requirements and acceptance criteria would be helpful e.g. in clarifying to what extent - if any - manufacturers that assay a large number of samples are penalized or rewarded, in regulatory practice it is often seen that requirements are expressed in terms of acceptance criteria for samples from the process or product, rather than in terms of product or process characteristics. Thus, the 1984 USP requirements to content uniformity of tablets [42] was formulated as the following acceptance criteria

Stage 1: Assay 10 tablets. Pass if both of the following criteria are met:

- 1) sample coefficient of variation is less than or equal to 6.0%
- 2) no value is outside claim $\pm 15\%$

Fail, if one or more values are outside claim $\pm 25\%$. Otherwise go to stage 2

Stage 2: Assay 20 further tablets. Pass if, for all 30 tablets, the following criteria are met:

- 1) sample coefficient of variation is less than or equal to 7.8 %
- 2) no more than one value is outside claim $\pm 15\%$, and no value is outside claim $\pm 25\%$.

Otherwise fail

Thus, in essence, the only requirement to product quality is that a (randomly selected) sample from the batch shall pass this test.

The requirements above have subsequently been subject to various amendments. The currently valid requirement, termed USP 24 [26], has been under revision since 1997. Although the various proposals tend towards a more parametric approach, the requirements are still formulated in terms of criteria to be applied to a randomly selected sample from the batch, and not as explicit requirements to content uniformity in the batch.

As considerations regarding sampling uncertainty are not explicit in these requirements, the implicit borderline between production that is considered satisfactory according to these requirements, and production that is not, is determined by the operating characteristics of these criteria. Therefore, in order for the manufacturer to design test and validation procedures and to make an informed choice of the number of samples to be used in such procedures, it is imperative to have an understanding of the operation of the criteria.

In the paper we shall therefore address some basic statistical issues related to a crude demonstration of uniformity, i.e. provision of evidence that there is a satisfactorily narrow dispersion of values in the entity under investigation.

The main issue of the paper is to discuss the influence of sampling uncertainty in such demonstrations of uniformity, and in particular to discuss the assurance provided against “unsatisfactorily large dispersions of values” under various

test procedures. Although acceptance criteria used in the pharmaceutical industry mostly are formulated in terms of requirements to sample results, we shall consider the operating characteristics of such procedures in terms of population values rather than properties of future samples from the population under investigation. Formulating the problem in terms of population values allows for interpretation in terms of the formal statistical framework of hypothesis testing and confidence intervals, and use of the concepts and techniques from the statistical theory of hypothesis testing to specify the assurance provided by the various procedures. In particular, concepts and techniques developed in the field of acceptance sampling may be used to provide insight in the mechanisms involved when providing assurance under due consideration to sampling uncertainty.

In the paper we shall discuss some of the acceptance criteria for blend or dose uniformity that have been suggested in the pharmaceutical literature. In Section 4 criteria based solely upon a measure of dispersion in the sample (standard deviation or coefficient of variation) are discussed. Such criteria are mainly used for blend uniformity analysis where use of statistical measures of location are not necessarily relevant (e.g. because of bias due to the sampling device).

The main body of the paper, Sections 5 to 6 discuss acceptance procedures that include a criterion on a measure of dispersion in the sample as well as criteria on individual measurement values like the two stage procedure in USP 21 [42]. It is shown that - in terms of population requirements - this procedure essentially monitors the proportion of population values outside some limiting values. The statistical problem of monitoring such a population value has been discussed in the literature on acceptance sampling by variables. In Section 5 this literature is reviewed and the assurance provided by various acceptance procedures based solely upon combinations of sample average and sample standard deviation that have been suggested in the pharmaceutical literature are discussed. Finally, in Section 6 the operating properties of the USP draft proposal [28] are discussed in light of these general results.

2 Acceptance criteria and statistical hypothesis testing

2.1 Choice of null hypothesis and alternative hypothesis

Whereas acceptance criteria are formulated as criteria to be applied to the assay results for a random sample from the blend or batch without reference to any distributional assumptions, the assessment of the operating properties of such criteria is usually best performed under specific assumptions on the distribution of sample values.

As the statistical theories of hypothesis testing provides a convenient formal statistical framework for design of acceptance criteria and assessment of discriminatory properties, the discussion of the various acceptance criteria will be related to concepts and results from statistical theories of hypothesis testing.

Thus, when a parametrized model for the sample results is assumed, we shall formulate a hypothesis (null hypothesis) H_0 regarding values of the parameters of the model (i.e. quality of the blend or batch). Those parameter values that do not belong to H_0 constitute the alternative hypothesis, H_1 .

It is an inherent feature of the Neyman-Pearson theory of hypothesis testing that a test can only offer evidence *against* the null hypothesis. A small observed significance, or p -value indicates that the alternative has significantly larger explanatory power. However, a large p -value does not suggest that the null hypothesis is true, but only that we lack evidence that it is not. Unfortunately, this difficulty is often swept under the carpet, the technically correct phrase “fail to reject the null hypothesis” being replaced by the term “accept”.

In general, when the goal of an experiment or sampling procedure is to establish an assertion, it is good statistical practice to take the negation of this assertion as the null hypothesis. The assertion becomes the alternative hypothesis. Therefore, when methods of statistical hypothesis testing are used for the purposes of conformity testing, i.e. to provide assurance that an entity conforms to a specified criterion, it is generally recommended to formulate the hypothesis

H_0 as the “nonconforming” parameter values, see also the discussion by Holst et al. [43]. Following this recommendation, the null hypothesis should be taken to mean “the batch is not satisfactory”, with the alternative being “the batch is satisfactory”. “Rejection” of the null hypothesis occurs when the sample has provided strong evidence favouring the assertion that the batch is satisfactory, and therefore, rejection of the null hypothesis should imply acceptance of the batch, whereas failure to reject the hypothesis is taken to mean that the sample has not provided sufficient evidence to warrant acceptance of the batch.

As this equivalence between “rejection of a hypothesis” and “acceptance of a batch” invariably leads to confusion, we have chosen in the paper when referring to statistical hypothesis testing to let the null hypothesis correspond to a region of “satisfactory population values”, and the alternative hypothesis correspond to unsatisfactory values. In this way the terms “accept” and “reject” may be used with the same meaning when referring to batch (or blend) acceptance criteria, and when referring to some underlying statistical hypothesis.

In the terminology of hypothesis testing, an acceptance criterion defines a region of potential sample results that do not satisfy this criterion, the *critical region*, such that whenever a sample result is within this region, the hypothesis is said to be rejected. The *significance level* of the test is the (maximum) probability of obtaining a sample result in the critical region when the hypothesis H_0 is true (i.e. the parameters of the distribution are as specified by the hypothesis).

As a consequence of this choice of null hypothesis to denote the values of population parameters that are “satisfactory” in some sense, it will sometimes be practical to use a critical region corresponding to a formal significance level of 95%, or 90%, say, instead of the traditional 5% or 10 % significance level.

Because of this duality in the choice of hypothesis and significance level it is advisable that the assessment of an acceptance criterion is performed by means of the *operating characteristic* of the criterion, i.e. the function showing the probability of passing the test as function of the parameters of the distribution of sample values (blend or batch quality) as this function shows the discriminatory power of the actual acceptance criterion, and does not otherwise depend

on phrasing of statistical hypothesis etc.

2.2 Confidence intervals and statistical tests

In practical assessment of sample results it is good statistical practice to consider not only the numerical estimate obtained from the sample, but also the associated uncertainty, e.g. in terms of a confidence interval giving a set of plausible values for the population parameter(s), i.e. a set of values that are in agreement (do not contradict) the sample result.

Sometimes acceptance criteria are formulated in terms of requirements that a $1 - \alpha$ confidence interval for the relevant parameter shall be fully included in some specified region of “satisfactory” qualities. Such an approach might be more easily understood than a criterion derived using formal statistical tests. There is, however, no conflict between these two approaches.

Thus, consider the formal statistical hypothesis concerning a population parameter, θ

$$H_0 : \theta > \theta_0 \text{ vs } H_a : \theta \leq \theta_0 \quad (1)$$

with θ_0 denoting some specified limiting value for the population parameter θ , and assume that some statistic T provides information about θ .

Then the rule: “Reject H_0 (and claim $\theta < \theta_0$) whenever the upper $(1 - \alpha)$ confidence limit for θ is less than the limiting value θ_0 ” corresponds to a level α test of the hypothesis H_0 , (or, alternatively to a level $(1 - \alpha)$ test of the hypothesis $H_0^* : \theta \leq \theta_0$). For a formal proof of the duality between statistical confidence regions and tests see e.g. [44] chapter 4.5.

The interpretation of any given acceptance criterion in terms of the critical region for some test is, however, more directly suitable for the derivation of the operating characteristic for that acceptance criterion. In particular it is useful when assessing the effect on the discriminatory power of using e.g. larger number of samples.

3 Notation and distributional assumptions

Following the usual practice in the pharmaceutical literature we shall not consider absolute dose values but assume that measurements are recorded relative to some (known) target dose weight or label claim.

As the focus of the paper is a discussion of the statistical properties of various acceptance criteria, we shall follow the practical convention in mathematical statistical literature and distinguish between random variables (i.e. the unknown sample result *before sampling*) denoted by capital letters, D_1, D_2, \dots, D_n , and actual values in a sample, d_1, d_2, \dots, d_n . Thus, D_1, \dots, D_n refers to a *probability distribution* of potential sample results, whereas d_1, \dots, d_n shall be understood as a generic representation of an actual set of sample results.

We shall let \bar{D} and S_d denote the average relative dose and standard deviation of relative doses in the sample,

$$\bar{D} = \sum_{i=1}^n D_i / n \quad (2)$$

$$S_d^2 = \sum_{i=1}^n (D_i - \bar{D})^2 / (n - 1) \quad (3)$$

with the corresponding sample results \bar{d} and s_d .

Moreover, as the term *relative sample standard deviation* may have different interpretations (as standard deviation of measurements relative to target value, or as standard deviation of measurements relative to the average sample value), we shall throughout the paper use the unambiguous term *sample coefficient of variation* to mean

$$Z = S_d / \bar{D}. \quad (4)$$

The discussion of the acceptance criteria and the derivation of expressions for acceptance probabilities is performed under the assumption that individual sample values, D_i may be represented by independent, identically distributed variables and that the distribution of sample results D_i may be described by a

normal distribution,

$$D_1, \dots, D_n \text{ mutually independent with } D_i \sim N(\mu_d, \sigma_d^2). \quad (5)$$

Thus, when D_i represents content of dosage units (tablets) selected from a batch, the assumption corresponds to assuming that the overall distribution of dose content in the batch may be represented by a normal distribution, and that individual dosage units are selected at random from the dosage units in the batch. When D_i represents content of blend samples, the assumption analogously corresponds to assuming that the overall distribution of such potential samples from the blend may be represented by a normal distribution, and that samples are taken at randomly selected positions in the blend.

To preserve the generality we shall use the term “population” when referring to this distribution of potential values representing dose content measurements from all doses in the tablet batch, or to the distribution of potential values representing conceivable blend sample measurements.

It should be noted that the model (5) will not be adequate when the distribution in the population is bimodal or multimodal corresponding e.g. to stratification between locations (e.g. top, middle and bottom), when the distribution is skewed, e.g. as a result of deblending, or when the distribution has heavier tails than the normal distribution, e.g. as a result of imperfect mixing (clustering) or of using dose particles that are too large for the intended dosage.

When sampling is performed from the “worst case positions in the blender”, as suggested e.g. in the FDA guidelines for blend analysis [20], the population refers only to sample results that might have been obtained under hypothetical repeated sampling from these positions, and the model does not necessarily reflect the population of values representing the totality of the blend.

When sampling is performed under a hierarchical (or nested) scheme as suggested among other by the Product Quality Research Institute, PQRI [18] where e.g. $n = 36$ sample values are obtained by repeatedly sampling from only 12 locations, the model will only be adequate in such (rather unlikely) situations where there is no correlation between subsamples from the same location in the blend or batch.

However, even despite these restrictions, the mathematical analytical discussion serves a purpose of clarifying and illustrating the statistical issues involved, thereby providing further insight in the properties of various tests that have been proposed in the pharmaceutical literature.

4 Acceptance criteria for the dispersion of doses

In pharmaceutical production, the blending, or mixing process is generally considered a key process aiming at producing a *uniform*, or *homogenous* blend. The purpose of blend sample analysis is to ensure that the blending is adequate for the end use.

As blend uniformity analysis might be subject to bias (i.e. systematic deviation of the mean value in the distribution of sample values) caused by the sampling, or by the analytical procedure, blend uniformity analysis puts main emphasis on assessment of the dispersion of population values.

4.1 Criterion based upon a specified limiting value of sample standard deviation

A direct estimate of the dispersion of dose-content is the standard deviation of the dose-content for the units in the sample. This estimate is an obvious quantity to use in making a judgment about the uniformity of a product. Further, the standard deviation is not affected by a constant bias of the mean due to sampling, handling (storage) or analytical disturbances.

Therefore, it might be considered to specify a limiting value, σ_{lim} for the sample standard deviation, and use the following acceptance criterion

Pass when the sample standard deviation, $s_d \leq \sigma_{lim}$ (6)
otherwise fail.

In the framework of statistical hypothesis testing, this acceptance criterion may

be interpreted as defining a critical region for a test of the hypothesis

$$H_0 : \sigma_d \leq \sigma_{lim} \text{ against the alternative } H_1 : \sigma_d > \sigma_{lim}. \quad (7)$$

The significance level of this test (6) of the hypothesis (7) is given as the probability of rejection in the borderline case when $\sigma_d = \sigma_{lim}$. In this borderline case there is approximately a 50% risk of obtaining a sample standard deviation exceeding σ_{lim} , and hence (referring to the hypothesis (7)) the significance level is approximately 50% regardless of the sample size.

A more detailed assessment of the operating characteristic for the acceptance criterion may be performed under specific assumptions for the distribution of sample values, D_i . Assuming a normal distribution of dosage units, (5), it follows that the sample variance S_d^2 is distributed like a $\sigma_d^2 \chi_{n-1}^2 / (n-1)$ distributed variable, and therefore the probability of acceptance may be determined from the χ^2 -distribution as

$$P_{acc}(\sigma_d) = P[S_d \leq \sigma_{lim}] = P \left[\chi_{n-1}^2 \leq (n-1) \left(\sigma_{lim} / \sigma_d \right)^2 \right] \quad (8)$$

i.e. from the cumulative distribution function of the χ_{n-1}^2 distribution. In the SAS-system software the acceptance probability may be found by calling the function `PROBCHI(x,df)` with the arguments $x = (n-1)(\sigma_{lim}/\sigma_d)^2$ and $df = n-1$ with σ_d denoting the (true) standard deviation between doses. In Microsoft Excel, the probability may be found from the function `CHIDIST(x,df)` providing a value which is one minus the probability of acceptance.

Figure 1 shows graphs of the probability of acceptance as function of the batch standard deviation of relative doses, σ_d , for $\sigma_{lim} = 0.06$ and various values of the sample size, n , between $n = 10$ and $n = 40$. It is seen that the set of curves tend to intersect at the point $\sigma_d = 0.06$, $P_{acc} = 0.5$ (the significance level), and that the discriminatory power (the steepness of the OC-curve) increases with increasing sample size.

When a sample size $n = 10$ is used, it is necessary that the batch standard deviation is less than 0.035 (producer's risk quality) in order to assure a high probability of acceptance, and the sampling plan does only provide a consumer protection (10% consumer risk) against a batch standard deviation of 0.085

(consumer's risk quality). As the limiting value of the sample standard deviation does not depend on the sample size, the consumer's and producer's risk qualities move towards this limiting value with increasing sample sizes, i.e. the (fixed) limiting value of the sample standard deviation, $\sigma_d = 0.06$ is also the borderline between batch standard deviations that are accepted under this criterion and batch standard deviations that are not accepted.

Moreover, it is a consequence of this choice of a limiting value for the sample standard deviation, independent of the size of the sample, that acceptance under this scheme provides a confidence of approximately 50% that the standard deviation in the blend (or batch) does not exceed the limiting value, $\sigma_{lim} = 0.06$, independent of the sample size (increasing from 45% for $n = 10$ to 48% for $n = 100$). However, when confidence statements corresponding to greater confidence are formulated, the assurance provided depends on the sample size. Thus, in the example considered, acceptance by a sampling plan with $n = 10$ provides a 90% confidence that the standard deviation in the blend does not exceed 0.09, whereas acceptance by a sampling plan with $n = 40$ provides a 90% confidence that the standard deviation in the blend does not exceed 0.07.

4.2 Criterion based upon a specified limiting value of sample coefficient of variation

In pharmaceutical applications, it is customary practice to use the sample coefficient of variation, (4) as a measure of dispersion.

This measure is used, eg. as part of the acceptance criterion in the test for uniformity of dosage units in the currently valid version of the United States Pharmacopeia, USP 24, [26].

It is interesting to note that although dosage values are already measured relative to the target, or label claim, and hence, the standard deviation of relative dosage values already measures dispersion in relative units, it has been considered appropriate, to make a further adjustment (for possible sampling or analysis bias) by considering the standard deviation relative to the estimate, \overline{D}

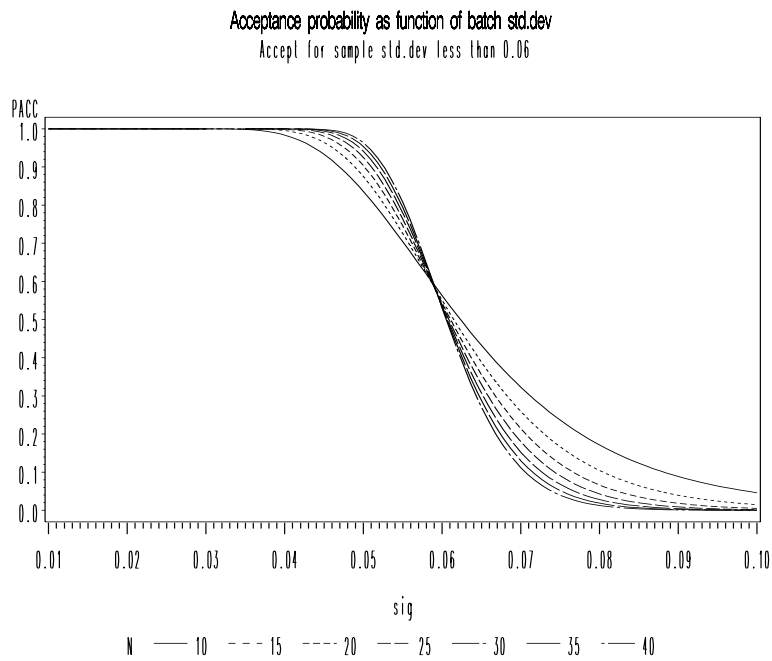


Fig. 1: The probability of accepting a batch as function of the relative standard deviation, σ_d for various sample sizes. Accept when sample standard deviation of relative doses is ≤ 0.06 .

of actual content.

Consider an acceptance criterion of the form

$$\begin{aligned} &\text{Pass when the sample coefficient of variation, } s_d/\bar{d} \leq C_{lim} \quad (9) \\ &\text{otherwise fail.} \end{aligned}$$

The acceptance criterion may be interpreted as defining a critical region for a test of the hypothesis

$$H_0 : \sigma_d/\mu_d \leq C_{lim} \text{ against the alternative } H_1 : \sigma_d/\mu_d > C_{lim}. \quad (10)$$

The operating characteristic for this acceptance criterion may be assessed under the assumption of a normal distribution of dosage units, (5) utilizing that the distribution of

$$T = \frac{\bar{D}}{S_d/\sqrt{n}} = \frac{\sqrt{n}}{Z} \quad (11)$$

is a non-central t-distribution with non-centrality parameter

$$\delta = \frac{\sqrt{n}\mu_d}{\sigma_d}$$

and $f = n - 1$ degrees of freedom (see e.g. Johnson and Kotz [45]) and hence for normally distributed measurements the distribution of the sample coefficient of variation depends only on the sample size and the batch coefficient of variation, $C_d = \sigma_d/\mu_d$.

Disregarding negative values of Z , it is seen that

$$P[Z \leq C_{lim}] = P\left[T \geq \frac{\sqrt{n}}{C_{lim}}\right] \quad (12)$$

and therefore the probability of acceptance may be found from the cumulative probability distribution of the noncentral t-distribution as

$$P_{acc}(C_d) = P[Z \leq C_{lim}] = 1 - P[t_{n-1}(\sqrt{n}/C_d) \leq \sqrt{n}/C_{lim}] \quad (13)$$

with $t_{n-1}(\sqrt{n}/C_d)$ denoting a random variable distributed according to a non-central t-distribution with $n-1$ degrees of freedom and noncentrality parameter \sqrt{n}/C_d .

In the SAS-system software the acceptance probability may be found by calling the function `PROBT(x,df,nc)` with the arguments $x = \sqrt{n}/C_{lim}$, $df = n - 1$ and $nc = \sqrt{n}/C_d$. This gives the probability of non-compliance; the probability of acceptance is then found as one minus the result provided by the function call.

Figure 2 shows graphs of the probability of acceptance as function of the batch coefficient of variation of relative doses, $C_d = \sigma_d/\mu_d$ for $C_{lim} = 0.06$ and various values of the sample size, n , between $n = 10$ and $n = 40$.

In analogy with the test of the sample standard deviation, Figure 1, the set of curves tend to intersect at the point $C_d = 0.06$, $P_{acc} = 0.5$ indicating that acceptance under this scheme provides a confidence of approximately 50% that the coefficient of variation in the blend (or batch) does not exceed the limiting value, $C_{lim} = 0.06$, independent of the sample size.

The difference between the acceptance probability corresponding to the criterion (7) on the sample standard deviation (relative to label claim) shown in Figure 1, and the acceptance probability corresponding to the criterion (10) on the sample coefficient of variation shown in Figure 2 is only marginal as long as the batch mean, μ_d does not deviate too much from the target value.

However, because of the sampling variation of the average dose, \bar{D} , that is introduced in the criterion (10) on the sample coefficient of variation, the curves corresponding to this criterion (Figure 2) are not quite as steep as their counterparts for the criterion on the sample standard deviation (Figure 1). Thus, when the coefficient of variation of dosage units in the batch is less than the limiting value, $C_{lim} = 0.06$, the acceptance probability under the sample coefficient of variation criterion is slightly less than under the sample relative standard deviation criterion, whereas it is slightly larger when the coefficient of variation in the batch is larger than 0.06.

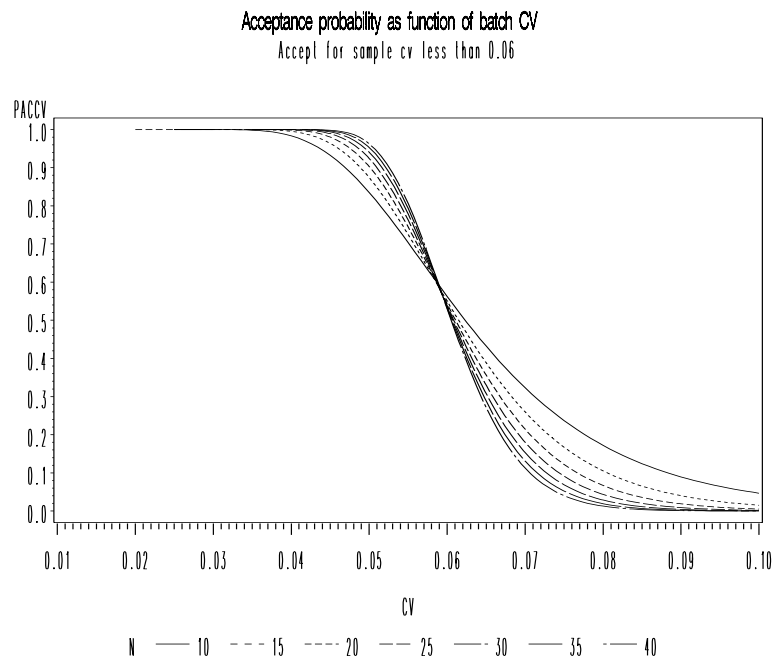


Fig. 2: The probability of accepting a batch as function of the coefficient of variation, $C_d = \sigma_d/\mu_d$ of doses in the batch for various sample sizes. Accept when sample coefficient of variation is ≤ 0.06 .

4.3 Criterion based upon prediction of standard deviation of future samples

When the acceptance criterion specifies a limiting value for some sample statistic, independent of the size of the sample, this limiting value implicitly defines a border value between population values that will pass under application of this criterion, and those that will not. For large samples, blends or batches with a population value that does not exceed this limiting value will pass under this criterion. For smaller samples the sampling uncertainty implies a lesser discriminatory power, and the amount of sample evidence favouring the hypothesis is given the same weight as the evidence against the hypothesis (the statistical test is performed at a 50% significance level).

This feature is not very satisfactory when the limiting value used in the acceptance criterion has been chosen to represent some undesirable quality, and when the purpose of the test is to provide some assurance against accepting batches of such qualities, and therefore the sample standard deviation should in some way be reduced by its uncertainty before comparison with the limiting value.

Such an acceptance procedure that puts more focus upon the assurance provided by the procedure has been suggested in PDA Technical Report 25 [22].

The fundamental idea in this procedure is that the sampling uncertainty associated with the sample under investigation shall be taken into account when deciding whether or not to accept the blend or batch. The blend is passed only if the sample provides strong evidence (even when accounting for the sampling uncertainty) that the quality is better than some prescribed undesirable quality.

The statistical technical tool invoked in the design of this procedure is the so-called *Standard Deviation Prediction Interval* to be explained below.

The starting point in the prediction interval approach as developed in [22] is a requirement (by authorities) that the standard deviation in a sample from the blend or batch shall not exceed a specified value, s_{spec} . Taking this to be the “specification” of satisfactory batch quality, the acceptance rule is then devised

to provide a high assurance that batches that are accepted by the sampling plan really are of a satisfactory quality. In [22] “satisfactory quality” is taken to mean that the standard deviation in a sample (of size 10) from the batch does not exceed $s_{spec} = 0.06$, and the “assurance” is chosen to be 90%.

It is worth noting that the “specification” of batch quality is not explicitly stated in terms of parameters like σ_d characterising the batch, but rather in terms of properties of a future sample from the batch.

Standard Deviation Prediction Interval

The standard deviation prediction interval, as described e.g. by Hahn and Meeker [46] is derived from the joint distributional properties of two sets of samples from the same population of normally distributed values, like two sets of samples, D_1, D_2, \dots, D_n (the current, actual sample) and D'_1, D'_2, \dots, D'_m (a future sample).

Let

$$S_d^2 = \sum_{i=1}^n (D_i - \bar{D})^2 / (n - 1) \quad (14)$$

denote the sample variance found in the current sample (of size n), and let moreover

$$S_p^2 = \sum_{i=1}^m (D'_i - \bar{D}')^2 / (m - 1) \quad (15)$$

denote the sample variance in a future sample of size m . Then, under the distributional model (5) assuming independence between the two sets of samples and assuming the same underlying population variance, σ_d^2 , for both samples, it follows that the distribution of

$$F = \frac{S_p^2}{S_d^2} \quad (16)$$

is a F-distribution with $(m - 1, n - 1)$ degrees of freedom, and hence

$$P \left[\frac{S_p^2}{S_d^2} \leq F(m - 1, n - 1)_{1-\alpha} \right] = 1 - \alpha \quad (17)$$

or, equivalently

$$P [S_p^2 \leq S_d^2 F(m - 1, n - 1)_{1-\alpha}] = 1 - \alpha \quad (18)$$

with $F(m - 1, n - 1)_{1-\alpha}$ denoting the $1 - \alpha$ quantile in the $F(m - 1, n - 1)$ distribution. The distribution does not depend on the actual population value σ_d^2 neither does it depend on the mean content in the population, μ_d (or a potentially different mean $\mu_{d'}$ for the future sampling process).

The reader is cautioned that in some (mainly US) textbooks and software quantiles and percentiles of non-normal statistical distributions are sometimes given as the so-called *critical* values indexed by the probability associated with more extreme values in the distribution. To avoid the ambiguity when using this practice in situations where one-sided as well as two-sided tails may be appropriate, we have chosen the unambiguous approach to let symbols like $F(m - 1, n - 1)_{1-\alpha}$, $\chi_{f,1-\alpha}^2$, $t_{f,1-\alpha}$, etc. refer to *quantiles* or *percentiles* of the distributions under consideration.

In the SAS-system software, the quantile may be found by calling the function `FINV(p,ndf,ddf)` with the arguments $p = 1 - \alpha$, $ndf = m - 1$, $ddf = n - 1$, and in Microsoft Excel, by calling the function `FINV(p,df1,df2)` with the arguments $p = \alpha$, $df1 = m - 1$, $df2 = n - 1$ (as the Microsoft Excel function returns the *critical value*).

Thus, when a value, $s_d = \sqrt{s_d^2}$, of the sample standard deviation is found in the sample, the upper $1 - \alpha$ prediction bound for the sample standard deviation in a future sample from that batch may be found as

$$s_{pred}(s_d) = s_d \sqrt{F(m - 1, n - 1)_{1-\alpha}}. \quad (19)$$

It follows from the formal calculations that as well the sampling uncertainty associated with the standard deviation in the actual sample, S_d , as the uncertainty associated with the standard deviation in the future sample, S_p , has been taken into account in the derivation of this upper bound.

Acceptance criterion

Now, assume that a limiting value, s_{spec} for the sample standard deviation in a future sample of size m has been specified.

Following the approach in [22] the following acceptance criterion may be considered.

$$\text{Pass when the standard deviation prediction bound, } s_{pred}(s_d) \leq s_{spec} \quad (20)$$

otherwise fail.

Utilizing the relation (19), and observing that $F(m-1, n-1)_{1-\alpha} = 1/F(n-1, m-1)_\alpha$ the acceptance criterion may be phrased in terms of a limiting value, s_{lim} of the sample standard deviation, s_d in the current sample,

$$s_{lim}(n) = s_{spec} \sqrt{F(n-1, m-1)_\alpha}. \quad (21)$$

In terms of this limiting value, the acceptance criterion is

$$\text{Pass when the sample standard deviation, } s_d \leq s_{lim}(n) \quad (22)$$

otherwise fail.

Clearly, the limiting value of the sample standard deviation depends on the sample size, n .

It may be shown that when the sample size, n , increases, the limiting value of the sample standard deviation, $s_{lim}(n) \rightarrow \sigma_{lim}$ with

$$\sigma_{lim} = s_{spec} \sqrt{(m-1)/\chi_{m-1, 1-\alpha}^2} \quad (23)$$

with $\chi_{m-1, 1-\alpha}^2$ denoting the $1-\alpha$ quantile in the χ^2 distribution with $m-1$ degrees of freedom. Thus, for $\alpha < 0.50$ the limiting value, $s_{lim}(n)$ of the sample standard deviation will be less than the specified value, s_{spec} , for future samples.

n							
10	15	20	25	30	35	40	∞
3.841	4.119	4.260	4.346	4.440	4.444	4.476	4.697

Table 1: Limiting value, $s_{lim}(n)$ of the sample standard deviation (in percent of target value) ensuring that the 90% prediction bound for a future sample of size 10 does not exceed 0.06

Table 1 shows values of the limiting value $s_{lim}(n)$ for various sample sizes, $n = 10, 15, \dots, 40$ for $s_{spec} = 0.06$ and $\alpha = 0.10$.

Under the distributional assumptions (5), the distribution of S_d , and, hence also of $s_{pred}(S_d)$ depends on the population standard deviation, σ_d , and therefore the acceptance criterion may be interpreted as defining a critical region for a test of the hypothesis

$$H_0 : \sigma_d \leq \sigma_{lim} \text{ against the alternative } H_1 : \sigma_d > \sigma_{lim} \quad (24)$$

with σ_{lim} given by (23). This limiting value of the population standard deviation depends on the number, m , of prospective sample items, and the desired confidence, $1 - \alpha$, that the standard deviation in this prospective sample shall be less than the design value, s_{spec} .

Operating characteristic

The operating characteristic for this acceptance criterion is given as

$$P_{acc}(\sigma_d) = \mathbb{P}[S_d \leq s_{lim}(n)] = \mathbb{P}\left[S_d \leq s_{spec} \sqrt{F(n-1, m-1)_\alpha}\right]. \quad (25)$$

Under the assumption of a normal distribution of measurement values, (5), values of the operating characteristic may be calculated utilizing that S_d^2 is distributed like a $\sigma_d^2 \chi_{n-1}^2 / (n-1)$ distributed variable, and therefore the probability of acceptance may be determined from the χ^2 -distribution as

$$P_{acc}(\sigma_d) = \mathbb{P}\left[\chi_{n-1}^2 \leq (n-1) \left(s_{lim}/\sigma_d\right)^2\right] \quad (26)$$

with s_{lim} given by (21), i.e.

$$P_{acc}(\sigma_d) = P \left[\chi_{n-1}^2 \leq (n-1) \left(s_{spec}/\sigma_d \right)^2 F(n-1, m-1)_\alpha \right]. \quad (27)$$

In the SAS-system software the acceptance probability may be found by calling the function **PROBCHI(x,df)** with the arguments $x = (n-1)(s_{spec}/\sigma_d)^2 F(n-1, m-1)_\alpha$ and $df = n-1$, and in Microsoft Excel, invoking the function **CHIDIST(x,df)** with the arguments $x = (n-1)(s_{spec}/\sigma_d)^2 F(n-1, m-1)_\alpha$ and $df = n-1$ returns a value which equals one minus the probability of acceptance.

Figure 3 shows graphs of the probability of acceptance as function of the population standard deviation of relative doses, σ_d , for various sizes, n , of the sample used for testing for the test with $s_{spec} = 0.06$, $m = 10$ and $\alpha = 0.10$, i.e. an assurance of 90% that the standard deviation in a future sample of size 10 does not exceed 0.06.

Comparing these graphs with the acceptance probability corresponding to a crude comparison of the actual sample standard deviation with the limiting value $\sigma_{lim} = 0.06$ shown in Figure 1, it is seen that the steepness of the curves remains unchanged, but the set of curves has been shifted to the left. Thus, the borderline between those batch standard deviations that will be accepted according to the SDPI-criterion, and those that are not accepted.

Under the standard deviation prediction interval acceptance criterion the probability of acceptance when $\sigma_d = 0.06$ is between 7% (for $n = 10$) and 1% (for $n = 40$).

When $\sigma_d = \sigma_{lim} = 0.047$ the probability of acceptance is around 20%. Thus, for the sample sizes under consideration, the significance level of the statistical test of the hypothesis H_0 formulated in (24) is 80%. In other words, the acceptance rule provides an assurance of 80% that the standard deviation in the population does not exceed $\sigma_{lim} = 0.047$.

In general, the significance level of the test of the hypothesis (24) performed using an acceptance criterion of the form (20) is found as $1 - P_{acc}(\sigma_{lim})$ with

$P_{acc}(\cdot)$ given by (27), i.e.

$$\alpha^* = 1 - \text{P} \left[\chi_{n-1}^2 \leq \frac{n-1}{m-1} \chi_{m-1, 1-\alpha}^2 \text{F}(n-1, m-1)_\alpha \right] \quad (28)$$

The significance level depends on the sample size. For increasing sample size, n , the significance level $\alpha^* \rightarrow 0.5$.

The borderline between those population standard deviations that will lead to acceptance according to this criterion, and those that are not accepted is $\sigma_d = 0.047$. Thus, for a sample size $n = 10$, it is necessary that the batch standard deviation is less than 0.03 (producer's risk quality) in order to assure a high probability of acceptance, but the sampling plan provides a consumer protection (10% consumer risk) against a batch standard deviation of 0.06 (consumer's risk quality). Using a larger sample, e.g. $n = 35$ allows the manufacturer to operate with a population standard deviation, σ_d up to 0.04 and still retain a high probability of acceptance, and, moreover, for this sample size, a batch with a standard deviation exceeding $\sigma_d = 0.05$, say, will only be accepted with a probability less than 15%.

4.4 A direct approach in terms of population values

Realizing that the approach based upon the standard deviation prediction interval is basically a statistical test of a hypothesis (24) concerning the population standard deviation σ_d , a more direct approach using concepts from statistical theories of hypothesis testing, one might choose a more direct approach and use the standard test for the hypothesis

$$H_0^* : \sigma_d \leq \sigma_{lim} \text{ vs } H_a^* : \sigma_d > \sigma_{lim} \quad (29)$$

with some specified limiting value, σ_{lim} , e.g. given by (23). The test may be found in introductory statistical textbooks (e.g. [47], sec. 10.13), and is also provided in the ISO Standard ISO 2854 [48] with the corresponding power curves given in ISO 3494 [49].

The acceptance criterion corresponding to a test of the null hypothesis (29) at

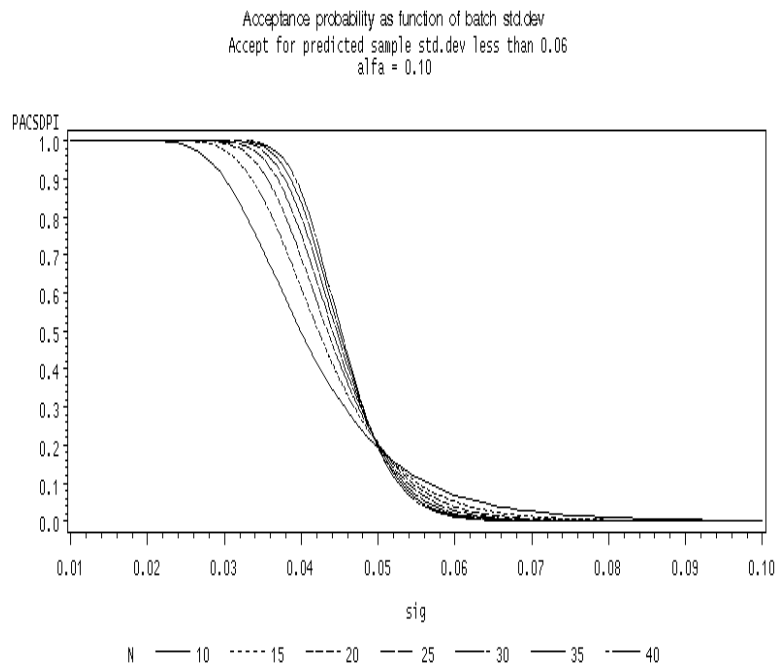


Fig. 3: The probability of accepting a batch as function of the relative standard deviation, σ_d for various sample sizes. Accept when standard deviation 90% prediction bound for a sample of size $m = 10$ is ≤ 0.06 .

significance level α^* is

Accept H_0^* when $s_d \leq s_{lim}(n)$
otherwise reject

with the limiting value, $s_{lim}(n)$, for the sample standard deviation given by

$$s_{lim}(n) = \sigma_{lim} \sqrt{\chi_{n-1, 1-\alpha^*}^2 / (n-1)} \quad (30)$$

and with $\chi_{f,p}^2$ denoting the p -quantile in the χ_f^2 -distribution.

As the purpose of the procedure is to provide assurance (even under consideration of the sampling uncertainty) that the value of σ_d is adequate, i.e. $\sigma_d \leq \sigma_{lim}$, the significance level of this test should be chosen rather high, e.g. a significance level, $\alpha^* = 0.8$

It follows from the discussion in Section 2.2 that the acceptance rule of this test is equivalent with an acceptance rule of the form “accept when the upper α^* confidence limit for σ_d (as calculated from s_d) does not exceed σ_{lim} . Thus, whenever a sample passes this test, there is a confidence of α^* (e.g. $\alpha^* = 0.8$ corresponding to 80% confidence) that the population standard deviation does not exceed σ_{lim} .

As the acceptance rule is intended to provide evidence that $\sigma_d \leq \sigma_{lim}$, one might have chosen (in line with good statistical practice, and the provisions in ISO 1576-1, [41]) instead to formulate a null hypothesis of the form

$$H_0 : \sigma_d > \sigma_{lim} \text{ vs } H_a : \sigma_d \leq \sigma_{lim} \quad (31)$$

with the interpretation that the content uniformity test is passed whenever the sample leads to rejection of this hypothesis (31) at a significance level, $\alpha = 0.20$, say.

The formulation in terms of a statistical test of this null hypothesis (31) might be more easily understood, as rejection of the hypothesis (31) takes place whenever the sample result provides strong evidence (at the magnitude $1 - \alpha$) against the hypothesis $\sigma_d > \sigma_{lim}$ in favour of the alternative $\sigma_d \leq \sigma_{lim}$, and, hence failure to reject the hypothesis $\sigma_d > \sigma_{lim}$ means that there is not sufficient evidence (considering the sample uncertainty) to claim that $\sigma_d \leq \sigma_{lim}$.

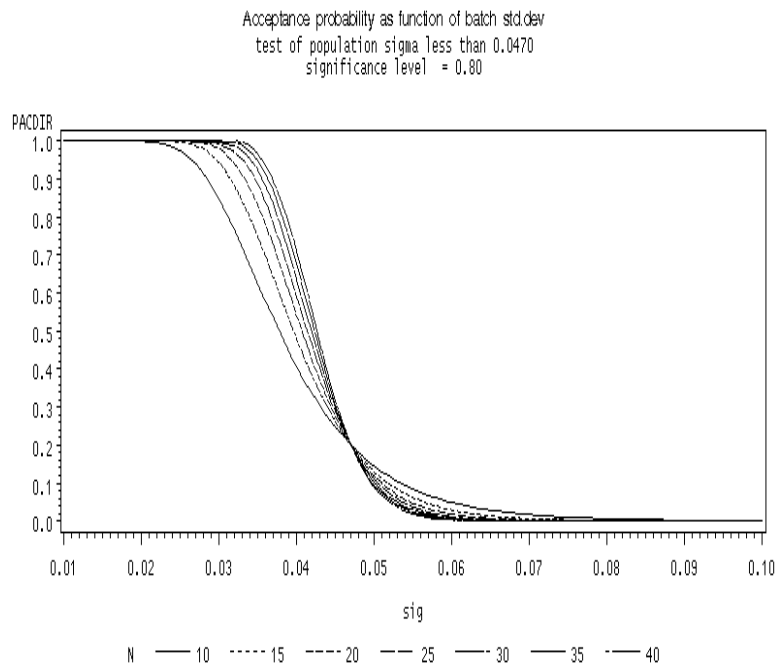


Fig. 4: The probability of accepting a batch as function of the relative standard deviation, σ_d for various sample sizes. Test the hypothesis $\sigma_d \leq 0.0470$ at significance level 0.80.

However, in order to maintain equivalence between the phrasing “passing the test” and “accepting the hypothesis” the statistical hypothesis to be tested has been formulated as in (29) in accordance with the discussion in Section 2.1.

5 Acceptance criteria with limits on individual measurements

5.1 The USP 21 criteria

Traditionally, tests for content uniformity of final product (dosage units) have not only been concerned with the dispersion of dosage values, but requirements to dispersion (e.g. to the sample coefficient of variation) have been supplemented by requirements on individual measurements.

Thus, in USP 21 [42] on each stage the criterion on the sample coefficient of variation is combined with criteria on individual measurements.

The requirements for passing this test for content uniformity are

Stage 1: Assay 10 tablets. Pass if both of the following criteria are met:

- 1) sample coefficient of variation is less than or equal to 6.0%
- 2) no value is outside claim $\pm 15\%$.

Fail, if one or more values are outside claim $\pm 25\%$. Otherwise go to stage 2

Stage 2: Assay 20 further tablets. Pass if, for all 30 tablets, the following criteria are met:

- 1) sample coefficient of variation is less than or equal to 7.8 %
- 2) no more than one value is outside claim $\pm 15\%$, and no value is outside claim $\pm 25\%$.

Otherwise fail.

From a statistical point of view two different issues are involved in these requirements

- a) In each stage, the criteria include as well a limit on the sample coefficient of variation (inspection by variables) as limits on individual measurements (inspection by attributes).
- b) The plan is a two-stage plan with a decision on whether to invoke the second stage depends on the result in the first stage, and the criteria to be applied in the second stage depends on the results in the combined sample (30 tablets).

The formulation of an acceptance sampling plan as a two-stage plan is a well-known principle in statistical theories of acceptance sampling. Most often a suitably designed two-stage plan allowing for acceptance as well as rejection in the first stage is developed as an alternative to an existing one-stage plan in order to provide the possibility of reaching a decision with less inspection effort, but without losing discriminatory power when compared with the one-stage plan. Use of a two-stage plan also has the psychological advantage that batches with borderline sample results in the first sample are given a second chance. For a discussion of two-stage sampling plans see Schilling, [12]. However, as the criteria invoked at the second stage involve the sample results from the first stage, the assessment of the properties of two-stage plans by analytical methods is not as straightforward as the assessment of one-stage plans. Therefore, in order to obtain insight in the statistical properties of the procedures, the assurance provided, and the interplay between the limits on the sample coefficient of variation (inspection by variables) and the limits on individual values (inspection by attributes), in the following we shall initially consider only one-stage procedures. In Section 6.2 we shall return to a discussion of the interplay between the two stages.

As the literature on assessment of the operating characteristics of acceptance procedures for dosage/blend uniformity mainly assumes a normal distribution of dosage units, the theoretical discussion will assume that the distribution of dosage values may be described by a normal distribution, and that sample measurements may be considered to be independent, identically distributed

according to this normal distribution (5).

Operating characteristic of the USP-21 test

Assuming a normal distribution of relative dosage units, $d_i \sim N(\mu, \sigma_d^2)$, and independent samples the probability of passing the USP-21 test is a function, $P_{usp}(\mu_d, \sigma_d)$, of the mean and standard deviation (μ_d, σ_d) of the dosage units in the population, and in principle, this function may be expressed by an analytical expression involving integration over a region in a 30-dimensional space of sample results.

It is well-known from statistical theory (see e.g. [44]) that under the assumption of independent, normally distributed sample values the sample average \bar{D} and sample standard variance S_d^2 are jointly sufficient for the population mean and variance, (μ_d, σ_d^2) , and therefore, when sample average and sample standard deviation (\bar{D}, S_d) are known, the distribution of individual measurement values (D_1, \dots, D_n) does only depend on these sample statistics, and not upon population mean and variance (μ_d, σ_d^2) . Thus, it is formally possible to devise a test based solely upon sample average and sample standard deviation in the two stages (otherwise disregarding individual measurements) having the same probability of passing as the USP-21 test for all values of population mean and standard deviation. However, as the level curves of the probability, $P_{usp}(\mu_d, \sigma_d)$ of passing the USP-test do not allow for a simple analytical representation, the actual mathematical derivation of such a test is not straightforward, and therefore more insight is gained by considering various approximations to $P_{usp}(\mu_d, \sigma_d)$.

By disregarding the correlation between the sample coefficient of variation and the individual measurements, Bergum [50] developed a simple expression giving a lower bound on the probability, $P_{usp}(\mu_d, \sigma_d)$, of passing the USP-test: For this lower bound level curves in the (μ_d, σ_d) -plane giving combinations, (μ_d, σ_d) , that have the same lower bound for this probability were provided. Bergum, op.cit. moreover compared the calculated lower bounds with simulated values of $P_{usp}(\mu_d, \sigma_d)$ and found a reasonable agreement.

5.2 Three-class attributes and a parametric approach

Operating characteristic of a three-class attribute plan

In the following we shall use a more direct analytical approach by (initially) disregarding the requirement on the sample coefficient of variation in the USP-21 criteria, and consider the following modification of the requirements for passing the USP-test obtained by disregarding the requirements on the sample coefficient of variation:

Stage 1: Assay 10 tablets. Pass if no value is outside claim $\pm 15\%$ ($LC \pm \Delta$)¹. continue to stage 2 if no more than one value is outside claim $\pm 15\%$ ($LC \pm \Delta$), and no value is outside claim $\pm 25\%$ ($LC \pm \Delta_1$).

Stage 2: Assay 20 further tablets. Pass if, for all these 20 tablets, no value is outside claim $\pm 15\%$ ($LC \pm \Delta$).

Otherwise fail.

The two sets of limiting values, $LC \pm \Delta$ and $LC \pm \Delta_1$ serve to specify a classification of each individual measurement values into one of three mutually exclusive classes defined by the following three zones around the target value, LC, viz.

Green zone: dose values in the interval $LC \pm \Delta$.

Amber zone: dose values in either of the intervals $LC - \Delta_1 < D < LC - \Delta$, or $LC + \Delta < D < LC + \Delta_1$.

Red zone: dose values in either of the intervals $D < LC - \Delta_1$, or $LC + \Delta_1 < D$.

¹In this paper calculations and discussions are based on measurements relatively to label claim, LC. Therefore in the formulae in the rest of the article, LC is equal to 1 (100%)

To aid the intuition the zones have been labelled with the colours on traffic light signals.

In terms of the three zones introduced above the requirements are:

Stage 1: Assay 10 tablets. Pass if all values are in the “green zone”, continue to stage 2 if no more than one value is in the “amber zone” and no values in the “red zone”.

Stage 2: Assay 20 further tablets. Pass if, for all these 20 tablets, all values are in the “green zone”.

Otherwise fail.

For single sampling plans (i.e. acceptance sampling plans with only one stage of sampling), sampling plans involving such attribute criteria with three classes are often referred to as a “three-class” attribute sampling plans. Such sampling plans were introduced by Bray et al [51] inspired by applications of lot acceptance sampling in food microbiology, and subsequently recommended by the International Commision on Microbiological Specifications for Foods (ICMSF) of the International Union of Microbiological Societies [52]. In the terminology of three-class sampling plans, the “green” zone is often labelled “good”, or “in compliance with upper limit of GMP”, the “amber” labelled “marginal” and the “red” labelled “bad”. Three class attribute sampling plans (to be used for one-stage sampling) usually specify sample size, maximum number, c_m , of non-good (i.e. non-green) items in the sample and maximum number, c_M , of bad (i.e. red) items in the sample. In microbiological applications c_M is often taken to be zero.

For normally distributed measurement values, D_i , the probability of an individual measurement value falling in the “red”, and the probability of a value in either the “amber” or the “red” zone, respectively, are

$$p_r(\mu_d, \sigma_d) = p_{nonc}(\mu_d, \sigma_d; \Delta_1) \quad (32)$$

$$p_a(\mu_d, \sigma_d) = p_{nonc}(\mu_d, \sigma_d; \Delta) \quad (33)$$

with the generic function $p_{nonc}(\mu, \sigma; \Delta)$ given by

$$p_{nonc}(\mu, \sigma; \Delta) = \Phi\left(\frac{LC - \Delta - \mu}{\sigma}\right) + 1 - \Phi\left(\frac{LC + \Delta - \mu}{\sigma}\right) \quad (34)$$

Hence, for normally distributed measurement values, the probability of a batch passing under this three-class criterion is

$$\begin{aligned} P_{3-c}(\mu_d, \sigma_d) &= (1 - p_a)^{10} + 10(p_a - p_r) \times (1 - p_a)^{29} \\ &= (1 - p_a)^{10} \times \left(1 + 10(p_a - p_r)(1 - p_a)^{19}\right) \end{aligned} \quad (35)$$

with p_a denoting the probability of a measurement value outside the interval $LC \pm \Delta$ (e.g. 85 to 115% of label claim) given by (33), and p_r given by (32) denoting the probability of a measurement value outside the interval $LC \pm \Delta_1$ (e.g. 75 to 125% of label claim) such that $p_a - p_r$ gives the probability of a measurement value outside $LC \pm \Delta$ (85 and 115%), but between $LC \pm \Delta_1$ (75 to 125%). Clearly, this latter probability does not exceed p_a .

Disregarding the contribution from p_r in (35), one obtains an approximation to the probability of passing the test as

$$P_{3-c,appr}(\mu_d, \sigma_d) = (1 - p_a)^{10} \times \left(1 + 10p_a(1 - p_a)^{19}\right). \quad (36)$$

This probability is slightly larger than the value given by (35). For any value of $p_a(\mu_d, \sigma_d)$, the largest value of $p_r(\mu_d, \sigma_d)$ occurs when $\mu_d = LC$ and $\sigma_d = \Delta/z_{1-p_a/2}$. For values of Δ and Δ_1 used in pharmaceutical practice, the approximation error incurred when substituting (35) by (36), i.e. by disregarding p_r , is only marginal.

Figure 5 shows the maximal value of the difference, $P_{3-c,appr}(\mu_d, \sigma_d) - P_{3-c}(\mu_d, \sigma_d)$ as function of the the proportion, $p_a = p_{nonc}(\mu_d, \sigma_d; \Delta)$ for $\Delta = 15\%$ and $\Delta_1 = 25\%$. The largest potential difference occurs for $p_a = 0.09$ (i.e. 9% of the population values falling in the amber or the red zone) with a maximum approximation error, $P_{3-c,appr}(\mu_d, \sigma_d) - P_{3-c}(\mu_d, \sigma_d) = 0.003$, i.e. a difference in acceptance probability of 0.3 %.

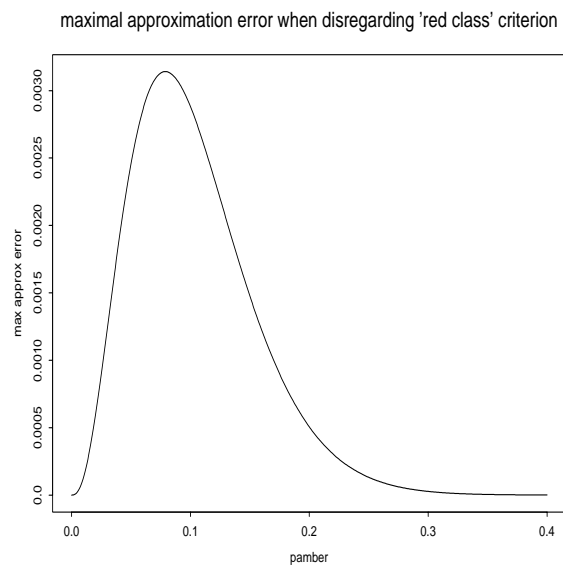


Fig. 5: Maximal absolute error in acceptance probability committed when disregarding the “red” class in calculation of acceptance probability versus total proportion of “amber” and “red” class units in the population. $\Delta = 0.15$ and $\Delta_1 = 0.25$.

P	$p_0(P)$
0.99	0.0070
0.95	0.0165
0.90	0.0244
0.50	0.0800

Table 2: Relation between acceptance probability, P , under the three class sampling plan, and limiting value, $p_0(P)$ of the population proportion of values outside $LC \pm \Delta$.

(Dahms and Hildebrandt [53] have discussed the choice of limits Δ and Δ_1 for applications in microbiological quality control under single sampling).

Population requirements derived from requirements to three-class acceptance probability

The approximative probability of passing the three-class attribute test, $P_{3-c,appr}$ is a decreasing function of p_a , and therefore a requirement

$$P_{3-c,appr}(\mu_d, \sigma_d) \geq P_{spec} \quad (37)$$

is equivalent with a requirement

$$p_a(\mu_d, \sigma_d) \leq p_0(P_{spec}) \quad (38)$$

with the limiting value, $p_0(P)$, of the population proportion nonconforming defined as the solution, p_0 , to

$$(1 - p_0)^{10} \left(1 + 10p_0(1 - p_0)^{19} \right) = P. \quad (39)$$

Table 2 shows the limiting value, $p_0(P)$, of the population proportion nonconforming corresponding to different values of the acceptance probability, P under the three class attribute criterion.

Hence, under the assumption of a normal distribution, the requirement of a specified probability of passing under this three-class test is equivalent to a

requirement that the proportion, $p_a(\mu_d, \sigma_d)$, of non-green values in the population does not exceed some limiting value that depends on the specified probability of passing. Thus, the apparent three class criterion (green, amber, red) is essentially just a two class criterion, (green, non-green).

In particular, a requirement of a 95% probability of passing under the three-class modification of the USP test is equivalent to requiring that no more than 1.65% of the units in the batch are outside the limits $LC \pm \Delta$ (85 and 115%).

Interpretation of a two-class requirement in terms of combinations of population mean and variance

It is well-known, see e.g. the textbook by Schilling [12] that for a specified fraction, p_0 , of nonconforming product the set of solutions (μ_d, σ_d) to

$$p_{nonc}(\mu_d, \sigma_d; \Delta) = p_0 \quad (40)$$

are found on a curve in the (μ_d, σ_d) -plane.

Thus, under the assumption of a normal distribution, control of the proportion of dosage units with content exceeding specified limiting values ($LC \pm \Delta$) is equivalent to controlling the combination of batch mean and variance.

Wallis [54] has provided a parametric description of the curve using an auxiliary variable, π , ($0 < \pi < p_0$) as

$$\mu_d = LC + \Delta \times (z_{p_0-\pi} - z_\pi) / (z_{p_0-\pi} + z_\pi) \quad (41)$$

$$\sigma_d = (\mu_d - LC - \Delta) / z_\pi \quad (42)$$

with z_p denoting the p -quantile in the standardized normal distribution. The idea is attributed by Wallis to Kenneth J. Arnold.

For $\pi = 0$ the point on the curve is $\mu_d = LC - \Delta$, $\sigma_d = 0$; for $\pi = p_0$ the point is $\mu_d = LC + \Delta$, $\sigma_d = 0$. When $\pi = p_0/2$, the parametric form gives $\mu_d = LC$, $\sigma_d = \Delta / z_{1-p_0/2}$ indicating that the maximum standard deviation in the population allowing for at most the fraction p_0 of product outside the limits $LC \pm \Delta$ is $\sigma_d = \Delta / z_{1-p_0/2}$ which is only admissible when the population mean is on target (LC).

Figure 6 shows examples of the curve (41), (42) for $\Delta = 16.5\%$ and for $p_0 = 0.0165$ and $p_0 = 0.0278$, respectively.

The curve is bounded by a trapezoid with corners in the points

$$\begin{aligned}
 A : & \quad \mu_d = LC - \Delta ; & \quad \sigma_d = 0 \\
 B : & \quad \mu_d = LC - \Delta(1 - z_{1-p_0}/z_{1-p_0/2}) ; & \quad \sigma_d = \Delta/z_{1-p_0/2} \\
 C : & \quad \mu_d = LC + \Delta(1 - z_{1-p_0}/z_{1-p_0/2}) ; & \quad \sigma_d = \Delta/z_{1-p_0/2} \\
 D : & \quad \mu_d = LC + \Delta ; & \quad \sigma_d = 0.
 \end{aligned} \tag{43}$$

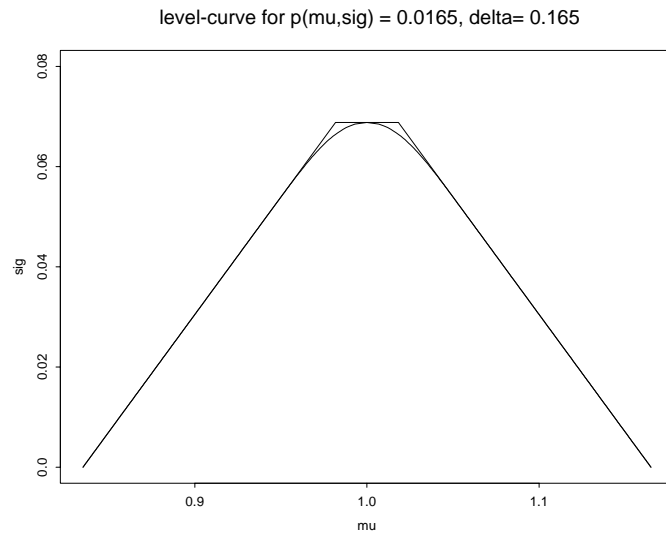
The length of the baseline is 2Δ , and the height of the trapezoid is $\Delta/z_{1-p_0/2}$. The slope of the left-hand side of the trapezoid is $1/z_{1-p_0}$. For $p_0 > 0.5$ the oblique sides of the trapezoid are sloping outwards, and the upper side of the trapezoid will be longer than the baseline. For $p_0 = 0.5$ the trapezoid degenerates to a rectangle, and for $p_0 < 0.5$ the oblique sides of the trapezoid are sloping inwards. The bounding trapezoids have also been shown in Figure 6.

Summary of properties of the three class attribute plan

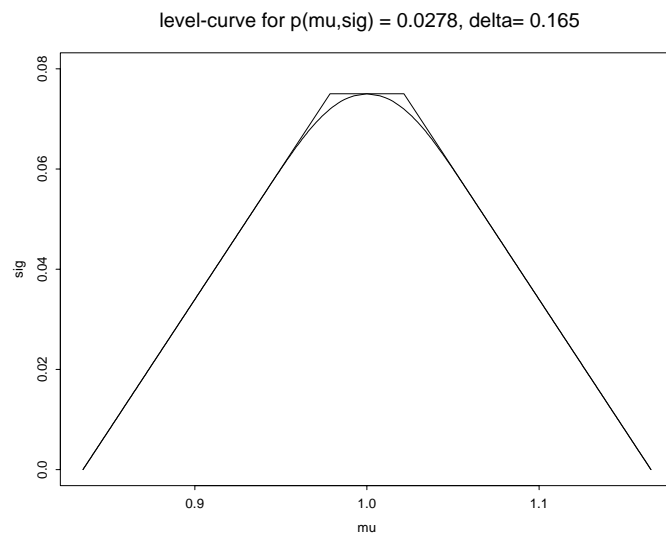
Disregarding the requirements to the sample coefficient of variation it was found that under a normal distribution of population values does the operating characteristic of the three class attribute criteria invoked at each stage in the USP 21 acceptance procedure only depend on the population mean and standard deviation through the proportion, $p_a(\mu_d, \sigma_d) = p_{nonc}(\mu_d, \sigma_d; \Delta)$, of units in the population with values outside the “green” zone ($LC \pm \Delta$).

Hence, assurance of a specified probability, P_{spec} , of acceptance under this three class attribute criterion is equivalent to assuring that the proportion, $p_{nonc}(\mu_d, \sigma_d; \Delta)$ of units in the population with values outside the interval $LC \pm \Delta$ does not exceed some value, $p_0(P_{spec})$, depending only upon the specified value, P_{spec} , of the acceptance probability.

In turn, assurance that the proportion, $p_{nonc}(\mu_d, \sigma_d; \Delta)$, of units in the population with values outside the interval $LC \pm \Delta$ does not exceed some specified value, p_0 , is equivalent to assuring that the combination (μ_d, σ_d) is within an



$$p_0 = 0.0165$$



$$p_0 = 0.0278$$

Fig. 6: Combinations of batch mean, μ_d , and standard deviation, σ_d , corresponding to a specified fraction, p_0 , of units outside the limits $LC \pm 16.5\%$.

(almost trapezoidal shaped) region in the (μ_d, σ_d) plane.

Hence, under the assumption of a normal distribution of population values, use of the three class attribute criterion is equivalent to a parametric requirement to the combination of population values (μ_d, σ_d) .

5.3 Confidence region approach

Consider a level curve for the probability, $P_{usp}(\mu_d, \sigma_d)$, of passing the USP-test under a normal distribution of dosage content, e.g. the curve corresponding to a probability γ of passing the test, i.e. values (μ_d, σ_d) satisfying

$$P_{usp}(\mu_d, \sigma_d) = \gamma \quad (44)$$

for that value of γ . The curve defines a region in the (μ_d, σ_d) -plane of combinations of batch mean and standard deviation with the properties that at least the proportion γ of all samples tested for content uniformity under the USP-test plan will pass the test.

Bergum [50] suggested a confidence region approach to construct acceptance limits for \bar{d} and s_d such that there is a specified assurance, $1 - \alpha$, that future samples from that batch will have a specified probability, γ , of passing the USP-test. The idea being that for each potential sample result, \bar{d} and s_d , a confidence region is constructed in the (μ_d, σ_d) -plane such that, for that sample result, there is a confidence of $(1 - \alpha)$ that the combination of batch mean and standard deviation is within this confidence region. Then, for each value, \bar{d} , of the sample mean in the interval $LC \pm \Delta$, a value of the sample standard deviation, $s_{d,cr}$, may be determined such that the point in the confidence region with the lowest probability of passing the USP-test is on the level curve corresponding to γ . The sample standard deviation determined in this way defines the acceptance limit for s_d corresponding to that value of the sample average, \bar{d} .

Thus, Bergum's approach shares the spirit of the prediction interval approach discussed in Section 4.3 in the sense that acceptance of the batch is based upon a strong confidence that a future sample from the batch will conform to

some specified requirements on that sample. In this way, when assessing the current sample, the uncertainty associated with the current sample is taken into consideration in order to provide the assurance desired.

PDA Technical Report 25 [22] provides tables of acceptance limits (expressed as the limiting value of the sample coefficient of variation, $s_{d,cr}/\bar{d}$ for each value of \bar{d}) for $n = 10$, $\gamma = 0.95$, and $1 - \alpha = 0.10$.

Analogously, the relations (37) and (39) might be used to formulate an acceptance criterion based upon a $(1 - \alpha)$ confidence interval for the proportion $p_a(\mu_d, \sigma_d)$ of doses in the batch with values in the amber and red zones, and accept the batch when the upper bound of this interval does not exceed $p_0(\gamma)$. Alternatively one could use a formulation of this criterion in terms of a statistical tolerance interval that contains at least the proportion $1 - p_0(\gamma)$ with confidence $(1 - \alpha)$, and accept the batch when the tolerance limits are within the green zone, see e.g. Hahn and Meeker [46].

5.4 Relation to theories of acceptance sampling by variables

The problem of designing sample-based acceptance criteria for monitoring the proportion, $p_{nonc}(\mu, \sigma; \Delta)$, of nonconforming product items (items outside the limits $LC \pm \Delta$) in a batch has been investigated in the literature on acceptance sampling. When the acceptance rule is based upon sample average and sample standard deviation, the term *acceptance sampling by variables for fraction nonconforming* is often used. In the following relevant results from that literature will therefore be summarized. When referring to the proportion of nonconforming units, we shall suppress the explicit reference to the limit Δ , and use the symbol $p_a(\mu_d, \sigma_d)$ to denote the proportion of units that do not conform to the limits $LC \pm \Delta$.

The decision whether to accept or reject a batch may be formulated in terms of

a test of the statistical hypothesis,

$$H_0 : p_a(\mu_d, \sigma_d) \leq p_0 \text{ with the alternative } H_a : p_a(\mu_d, \sigma_d) > p_0 \quad (45)$$

for a specified value of p_0 .

In order to maintain equivalence between the phrasing “passing the test” and “accepting the hypothesis” the statistical hypothesis to be tested in (45) has been formulated in accordance with the discussion in Section 2.1.

The statistical problem of testing a hypothesis of the form (45) was addressed by Wallis, [54], who suggested to use a test based upon the estimator of $p_a(\mu_d, \sigma_d)$ obtained by substituting μ_d and σ_d in the generic function (34) by the usual estimates, \bar{d} and s_d

$$\text{Accept } H_0 \text{ if } p_a(\bar{d}, s_d) \leq p_{lim}, \text{ otherwise reject} \quad (46)$$

i.e. use the sample average and sample standard deviation to compute an estimate of the fraction nonconforming items using (33), and compare this estimate to some limiting value, p_{lim} .

Subsequently, Lieberman and Resnikoff [33] suggested an approach based upon an optimal (UMVU) estimator of $p_a(\mu_d, \sigma_d)$. This approach was adapted in the US MIL STD 414, [55], and as this standard was solving a need in industry, this standard was later adopted as an ISO Standard, ISO 3951, [35], and as a US national standard ANSI Z1.9 [56], see also Schilling [57] and Boulanger et al. [58].

In light of more recent developments of theories of hypotheses testing, various authors have investigated other approaches than the MVU estimator-based approach suggested by Lieberman and Resnikoff, see e.g. Bruhn-Suhr et al. [59], Lei and Vardeman [60].

In the following we shall follow the direct approach suggested by Wallis [54]. The difference between this direct approach by Wallis and use of the likelihood-ratio principle is that under the likelihood-ratio approach, the estimate, s_d of σ_d is replaced by the maximum-likelihood estimate, $\left(\sqrt{(n-1)/n}\right)s_d$. For a comparison of the direct approach and the likelihood-ratio approach see Lei and Vardeman [60].

Although the problem is a rather common problem in industrial practice, tests of the hypothesis (45) are seldomly described in statistical textbooks.

Once, the limiting value, p_{lim} , of the estimated fraction nonconforming has been established, the acceptance region (in the (\bar{d}, s_d) -plane) for the test may be determined as the set of values, (\bar{d}, s_d) , such that

$$p_a(\bar{d}, s_d) \leq p_{lim} \quad (47)$$

with the function $p_a(\cdot, \cdot)$ given by (33) and (34). The acceptance region is of a similar shape as shown in Figure 6. The boundary of the region may be determined from (41) and (42) substituting p_{lim} for p_0 . When the significance level of the test is less than 50%, p_{lim} will be larger than p_0 (the sampling uncertainty is “subtracted” from the estimated proportion nonconforming before comparing with p_0), and the acceptance region will include the region corresponding to $p_a(\bar{d}, s_d) \leq p_0$. When the significance level of the test is greater than 50%, p_{lim} will be less than p_0 (the sampling uncertainty is “added” to the estimated proportion nonconforming before comparing with p_0), and the acceptance region will be inside the region corresponding to $p_a(\bar{d}, s_d) \leq p_0$.

The acceptance region may be expressed as a list of limiting values, $s_{lim}(\bar{d})$, for the sample standard deviation, s_d , with the limiting values depending on the sample average, \bar{d} . The maximal acceptable value of the sample standard deviation is found from (42) for $\pi = p_{lim}/2$, $\bar{d} = LC$ as

$$s_{max} = \Delta / z_{1-p_{lim}/2}. \quad (48)$$

Alternatively, the acceptance region may be expressed as a list of limiting values $CV_{lim}(\bar{d}) = s_{lim}(\bar{d})/\bar{d}$ for the sample coefficient of variation, with the limiting values depending on the sample average, \bar{d} and with the maximal acceptable sample coefficient of variation given by (48).

For any combination, (μ_d, σ_d) , of batch mean and standard deviation, the probability of passing this dosage uniformity test with a specified limiting estimated proportion nonconforming, p_{lim} , may be determined by numerical integration of the joint probability density function for (\bar{d}, s_d) over the acceptance region (in the (\bar{d}, s_d) -plane). It follows from the mutual statistical independence of \bar{d}

and s_d (see [44]) that the joint probability function may be determined as the product of density functions corresponding to the normal distribution of \bar{d} and the $\sigma_d^2 \chi^2(n-1)/(n-1)$ distribution of s_d^2 .

Hence, for any specified limiting value, p_{lim} , the probability of passing this test may be determined as function of population parameters (μ_d, σ_d) . As the test is based upon a comparison of the estimated fraction nonconforming items with some limiting value, it is to be expected that the probability of passing the test depends only on the proportion, $p_a(\mu_d, \sigma_d)$ of nonconforming items in the population. Although this is not exactly true, the assertion generally holds to a degree of approximation satisfactory for all practical purposes, see Lei and Vardeman [60].

The limiting value, p_{lim} , is determined from the requirement to the assurance desired from the test procedure. In statistical terms, the assurance is expressed in terms of the significance level, α , of the statistical test. Thus, using significance level, $\alpha = 0.90$ say, the borderline case with population proportion nonconforming, $p_y(\mu_d, \sigma_d) = p_0$, the probability of rejecting H_0 is 90% (the probability of rejecting H_0 is α).

When the design value p_0 in (45) has been selected as the value $p_0(\gamma)$ derived from (37) and (39), it follows from the duality between statistical confidence regions and tests that acceptance by the test at a significance level α provides assurance with confidence $1 - \alpha$ that there is at least a probability γ that a future sample from the batch will pass the three-class modification of the content uniformity test.

5.5 Design of a test with a trapezoidal acceptance region

The shape of the acceptance region in (\bar{d}, s_d) -space corresponding to a specified significance level, α , is more easily understood by first considering the derivation of a statistical test referring to a one-sided specification limit, e.g. a specified upper limit, $U = LC + \Delta$.

Test of a one-sided upper specification

Consider a test concerning a one-sided specification limit, e.g. an upper limit, U . The quantity of interest is

$$p^+(\mu_d, \sigma_d) = P[U < D] = 1 - \Phi\left(\frac{U - \mu_d}{\sigma_d}\right). \quad (49)$$

Thus, we shall consider the hypothesis

$$H_0^+ : p^+(\mu_d, \sigma_d) \leq p_0 \text{ vs } H_1^+ : p^+(\mu_d, \sigma_d) > p_0 \quad (50)$$

for some specified value p_0 and with $p^+(\cdot, \cdot)$ given by (49).

Now introduce the transformed random variable, $Z = U - D$. It follows that $Z \sim N(U - \mu_d, \sigma_d^2)$, and hence,

$$p^+(\mu_d, \sigma_d) = P[Z < 0] = \Phi\left(\frac{0 - (U - \mu_d)}{\sigma_d}\right) = \Phi(-\theta)$$

with $\theta = (U - \mu_d)/\sigma_d$.

In terms of θ we have $Z \sim N(\theta\sigma_d, \sigma_d^2)$.

Thus, there is a one-to-one correspondance between $p^+(\mu_d, \sigma_d)$ and θ , viz. $\theta = z_{1-p}$, corresponding to $\mu_d = U - \sigma z_{1-p}$.

Accordingly, the hypothesis (50) is equivalent with a hypothesis

$$H_{0,\theta}^+ : \theta \geq z_{1-p_0} \text{ vs } H_{1,\theta}^+ : \theta < z_{1-p_0}. \quad (51)$$

Let \bar{Z} and S_z^2 denote average and empirical variance of Z_1, Z_2, \dots, Z_n . It follows from a result by Lehmann [61] that the uniformly most powerful invariant test for $H_0^{+\theta}$ is a test with critical region of the form:

$$\text{Reject } H_{0,\theta}^+ \text{ for } \bar{z}/s_z < k \quad (52)$$

The significance level of this test is determined from the distribution of $W^+ = \bar{Z}/S_z = (U - \bar{d})/S_d$. A large value of W^+ indicates a small fraction in the population exceeding the upper limiting value U .

The distribution of

$$T = \sqrt{n} \frac{U - \bar{D}}{S_d} = \sqrt{n} W$$

is a non-central t-distribution with $n - 1$ degrees of freedom and with non-centrality parameter $\delta = \theta\sqrt{n}$.

As $W^+ = T/\sqrt{n}$ it follows that

$$P[W^+ \leq k] = P[T/\sqrt{n} \leq k] = P[T \leq k\sqrt{n}]$$

and, hence

$$P[(U - \bar{D})/s_d \leq k] = P[W^+ \leq k] = P[t(n - 1, \delta(p)) \leq k\sqrt{n}] \quad (53)$$

with $\delta(p) = \theta\sqrt{n} = z_{1-p}\sqrt{n}$.

Thus, the critical value, k (for w^+) at a level α test for the hypothesis (51) is determined from

$$k = k(p_0) = t(n - 1, \delta(p_0))_\alpha / \sqrt{n} \quad (54)$$

viz. the α percentile in the $t(n - 1, \delta(p_0))$ -distribution. The hypothesis (51) is accepted when $w^+ \geq k(p_0)$, and therefore, whenever $w^+ \geq k(p_0)$ it has been demonstrated with a confidence α that $p^+(\mu_d, \sigma_d) \leq p_0$.

For $p_0 = 0.0165$, $n = 10$ and $\alpha = 0.90$ one finds

$$k(0.0165) = 3.255.$$

The requirement determines a region in the (\bar{d}, s_d) -plane

$$\bar{z} \geq s_d k$$

or,

$$U - \bar{d} \geq s_d k.$$

Inserting $U = LC + \Delta$, the requirement $w^+ \geq k$ may be expressed as

$$LC + \Delta - \bar{d} \geq s_d k(p_0). \quad (55)$$

Thus, the acceptance criterion is of the form

$$\begin{aligned} &\text{Accept } H_0^+ \text{ when } \bar{d} + ks_d \leq \text{LC} + \Delta \\ &\text{otherwise reject} \end{aligned} \quad (56)$$

with $k = k(p_0)$ given by (54). In this form the criterion simply compares a suitably chosen linear combination of sample average and sample standard deviation with the upper specification for individual measurement values.

Graphically, the criterion, (55) determines a line

$$s_d = (\text{LC} + \Delta - \bar{d})/k \quad (57)$$

in the (\bar{d}, s_d) -plane, passing through the point $(\bar{d} = \text{LC} + \Delta, s_d = 0)$ with slope $-1/k$ for $k = k(p_0)$ given by (54).

For sample values, (\bar{d}, s_d) , to the left of this line the hypothesis (50) is accepted. It follows from (53) that the probability of obtaining a sample value, (\bar{d}, s_d) , to the left of this line depends on (μ_d, σ_d) only through $p^+(\mu_d, \sigma_d)$. In other words, all combinations of batch mean and standard deviation (μ_d, σ_d) with the same proportion, $p = p^+(\mu_d, \sigma_d)$ of nonconforming units (with respect to the upper limit $\text{LC} + \Delta$) will have the same probability of rejection under this test. In statistical terms the critical region is said to be “similar”.

Introducing the estimated fraction nonconforming units (with respect to the upper limit U),

$$\hat{p}^+ = p^+(\bar{d}, s_d) \quad (58)$$

it may be verified that the acceptance criterion (56) is equivalent with

$$\begin{aligned} &\text{Accept } H_0^+ \text{ when } p^+(\bar{d}, s_d) \leq p_{lim} \\ &\text{otherwise reject} \end{aligned} \quad (59)$$

with the limiting value, p_{lim} for the estimated fraction nonconforming determined from

$$p_{lim} = \Phi(-k) \quad (60)$$

with $k = k(p_0)$ given by (54). Thus, the limiting value for the estimated fraction nonconforming depends as well upon the sample size as upon the significance level, α , of the test (the desired level of assurance of $p \leq p_0$).

For a level α -test, the value of k is given by (54) and the acceptance probability for this test is determined from (53) with that value of k , i.e.

$$P[(U - \bar{D})/s_d \geq k] = P[t(n-1, \delta(p)) > t(n-1, \delta(p_0))_\alpha] \quad (61)$$

with $\delta(p) = \theta\sqrt{n} = z_{1-p}\sqrt{n}$.

In the SAS-system, the acceptance probability may be determined by means of the function `PROBT(T, DF, NCNTL)` as

$$PACC = 1 - \text{PROBT}(K, N-1, NCNTL)$$

with $K = \text{TINV}(\text{ALFA}, N-1, \text{PROBIT}(1-P0) * \text{SQRT}(N))$, and $NCNTL = \text{SQRT}(N) * \text{PROBIT}(1-P)$.

Test of a one-sided lower specification

Consider a test concerning the lower limit, $L = LC - \Delta$, with the quantity of interest

$$p^-(\mu_d, \sigma_d) = P[D < L] = \Phi\left(\frac{L - \mu_d}{\sigma_d}\right) \quad (62)$$

and the hypothesis

$$H_0^- : p^-(\mu_d, \sigma_d) \leq p_0 \text{ vs } H_1^- : p^-(\mu_d, \sigma_d) > p_0. \quad (63)$$

By symmetry, it is found that a level α -test of the hypothesis (63) rejects the hypothesis when

$$\bar{d} - LC - \Delta < s_d k \quad (64)$$

with $k = k(p_0)$ given by (54).

Accordingly, the acceptance criterion corresponding to a test at significance level α of the hypothesis (63) regarding the lower limit is

$$\text{Accept } H_0^- \text{ when } (LC - \bar{d}) + k s_d \leq \Delta \quad (65)$$

otherwise reject.

The criterion (65) determines a line

$$s_d = (\bar{d} - LC - \Delta)/k \quad (66)$$

in the (\bar{d}, s_d) -plane, passing through the point $(\bar{d} = LC - \Delta, s_d = 0)$, and with slope $1/k$ where $k = k(p_0)$ is given by (54).

Introducing the estimated fraction nonconforming units (with respect to the lower limit L)

$$\hat{p}^- = p^-(\bar{d}, s_d) \quad (67)$$

it is seen that the acceptance criterion (65) is equivalent with the criterion:

$$\begin{aligned} &\text{Accept } H_0^- \text{ when } p^-(\bar{d}, s_d) \leq p_{lim} \\ &\text{otherwise reject} \end{aligned} \quad (68)$$

with p_{lim} determined from (60) for $k = k(p_0)$ given by (54).

Combination of the two tests

So far, the upper and lower limits, U and L , have been considered separately. However, the quantity of interest is the proportion,

$$p_a(\mu_d, \sigma_d) = p^-(\mu_d, \sigma_d) + p^+(\mu_d, \sigma_d) \quad (69)$$

of units violating either of the limits.

Accordingly, the estimates \hat{p}^+ from (58) and \hat{p}^- from (67) may be combined to form an estimate,

$$\hat{p}_a(\bar{d}, s_d) = p^-(\bar{d}, s_d) + p^+(\bar{d}, s_d) \quad (70)$$

of the proportion of units that do not conform to the combined specification. It seems natural to base acceptance of the hypothesis (45) concerning $p_a(\mu_d, \sigma_d)$ upon a comparison of this estimated proportion of nonconforming units, $\hat{p}_a(\bar{d}, s_d)$ with some limiting value p_{lim} .

$$\begin{aligned} &\text{Accept } H_0 \text{ when } \hat{p}_a(\bar{d}, s_d) \leq p_{lim} \\ &\text{otherwise reject} \end{aligned} \quad (71)$$

where p_{lim} shall be determined to secure a significance level, α , of the test.

A parametric form of the boundary of the region in the (\bar{d}, s_d) -plane satisfying $\hat{p}_a(\bar{d}, s_d) \leq p_{lim}$ is given by (41) and (42).

For any combination, (μ_d, σ_d) of the population values, the acceptance probability under this criterion may be determined by numerical integration of the joint probability density function for (\bar{D}, S_d) over this region.

However, as noted by Lei and Vardeman [60] for any specified limiting value, p_{lim} , there is a whole band of operating characteristic curves, and there the determination of a limiting value p_{lim} corresponding to a specified significance level, α , of the test of the hypothesis (45) requires extensive use of numerical search methods.

Instead, following the suggestion by Wallis, one might construct an acceptance region corresponding to a level α test of the two-sided specification by combining the acceptance regions corresponding to level α -test of each of the two one-sided specifications, and supplement with a requirement to the maximum sample standard deviation derived from a level α -test of the two-sided specification. Following this suggestion, the acceptance region will be a trapezoidal region in (\bar{d}, s_d) -space with oblique sides given by the lines (66), (57) corresponding to that value of α , and with upper side given by

$$s = s_{max} = \Delta / z_{1-p^*/2} \quad (72)$$

with $p^* = \Phi(-k)$.

Thus, the resulting criterion is of the form

$$\begin{aligned} &\text{Accept } H_0 \text{ when } (\bar{d}, s_d) \text{ is within the polygon specified by } k \quad (73) \\ &\text{otherwise reject.} \end{aligned}$$

It follows from the considerations in Section 5.4 that the region in (\bar{d}, s_d) -space of values (\bar{d}, s_d) such that $p_a(\bar{d}, s_d) \leq p_{lim}$ is bounded by this acceptance polygon. Thus, the acceptance criterion (73) corresponding to the polygon will be slightly more conservative (i.e. accept in more cases) than the criterion

(71) based upon a direct comparison of $p_a(\bar{d}, s_d)$ with p_{lim} . However, as indicated in Figure 6, the difference between these two acceptance regions is only marginal. The difference is most apparent for larger values of s_d close to s_{max} . Hence, for any given value of the proportion nonconforming, $p_a(\mu_d, \sigma_d)$, the largest difference in acceptance probability occurs when $\mu_d = LC$. This maximal difference in probability of acceptance under the two criteria has been illustrated as function of $p_a(\mu_d, \sigma_d)$ in Figure 7. The largest difference occurs for acceptance probabilities around 0.50.

Alternative representations of the acceptance rule

The acceptance criterion (73) specifying a trapezoidal acceptance region in (\bar{d}, s_d) -space may be represented in various equivalent ways.

Rearranging (55), the requirement for acceptance at the upper limit may be expressed as

$$s_d \leq (LC + \Delta - \bar{d})/k$$

or,

$$s_d/\bar{d} \leq (LC + \Delta - \bar{d})/(k\bar{d}) \quad (74)$$

which, corresponding to each value of the sample average, \bar{d} , specifies an upper limit to the sample coefficient of variation to be applied for $LC < \bar{d}$.

Similarly, from rearranging (64), the requirement for acceptance at the lower limit may be expressed as

$$s_d \leq (\bar{d} - LC - \Delta)/k$$

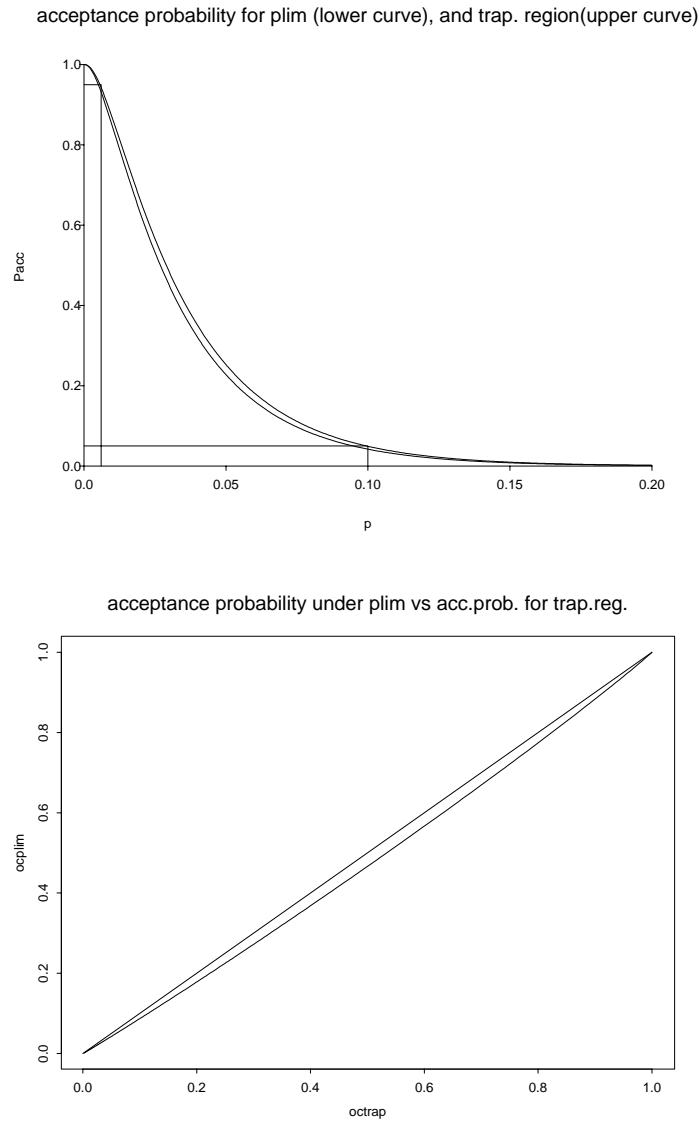
or,

$$s_d/\bar{d} \leq (\bar{d} - LC - \Delta)/(k\bar{d}) \quad (75)$$

providing an upper limit to the sample coefficient of variation to be applied for $\bar{d} < LC$.

For values of \bar{d} in the interval,

$$LC - \Delta(1 - z_{1-p_{lim}}/z_{1-p_{lim}/2}) \leq \bar{d} \leq LC + \Delta(1 - z_{1-p_{lim}}/z_{1-p_{lim}/2}) \quad (76)$$



$$\Delta = 0.165; n = 21, k = 1.91, s_{max} = 0.075; p_{lim} = 0.0278; \mu_d = LC.$$

Fig. 7: Top: Acceptance probability under the criterion (71) based upon the estimated proportion nonconforming, and under the criterion (73) for the corresponding trapezoidal rule. Bottom: $P - P$ -plot of corresponding values of the acceptance probability under the two criteria. The identity line has been superimposed on the graph.

corresponding to the center part of the trapezoid (see (43)), the criterion to the sample standard deviation is

$$s_d \leq s_{max}$$

with the maximum sample standard deviation, s_{max} given by (48). Hence, for these values, the criterion to the sample coefficient of variation is

$$s_d/\bar{d} \leq s_{max}/\bar{d}. \quad (77)$$

The requirements (74), (75) and (77) may be presented in a tabular form as in PDA Technical Report No. 25 [22].

Combining (56) and (65) it is seen that the region bounded by the oblique sides of the trapezoid may be delimited by the requirement

$$|LC - \bar{d}| + ks_d < \Delta \quad (78)$$

i.e. the absolute value of the deviation between sample average and LC plus (or minus, depending on the sign of k) some adjustment for sampling uncertainty shall not exceed the specification Δ (e.g. 16.5%). It follows from (54) that when the level of significance, α has been chosen to be larger than 0.50, then $k > 0$ and, hence the uncertainty adjustment is added as an extra penalty, whereas when $\alpha < 0.50$, the uncertainty adjustment is subtracted before comparison with Δ .

The adjustment for sampling uncertainty

Irrespective of the representation of the acceptance rule, considerations on the approach to the influence of the sampling uncertainty are reflected through the choice of significance level, α , that in turn is determinant for the value k given by (54) that is used in the acceptance rule.

In this section we shall briefly describe the effect of the choice of significance level on the designation of the various limiting values. For ease of transparency, we shall first consider a test of a one-sided upper specification as discussed in the subsection page 149.

From (55) it is seen that in essence the criterion determines an estimate of z_{1-p_0} as $(LC + \Delta - \bar{d})/s_d$, and compares this estimate with a limiting value k . If there was no sampling uncertainty, and \bar{d} and s_d represented the population values (μ_d, σ_d) then k should be chosen to be z_{1-p_0} . However, accounting for the sampling uncertainty (in accordance with the specified level of significance, α ,) the limiting value, k will also depend upon the level of significance, α , and the sample size, n , as seen from (54).

The approach to the adjustment for sampling uncertainty depends on the value of the significance level α . When $\alpha > 0.50$ (and $p_0 < 0.5$) then k will be larger than z_{1-p_0} , implying that the sample average shall be even less than

$$\bar{d}_{lim}(s_d) = (LC + \Delta) - z_{1-p_0}s_d \quad (79)$$

in order to provide satisfactorily strong evidence of $p_a < p_0$.

For significance levels $\alpha < 0.50$ the value of k will be smaller than z_{1-p_0} , implying that even sample averages exceeding $\bar{d}_{lim}(s_d)$ may lead to acceptance, if only the exceedance is not larger than what may be attributed to the sampling uncertainty. Thus, for $\alpha < 0.50$ the procedure accepts unless there is strong sample evidence *contradicting* $p_a < p_0$.

The larger the sample, the smaller is the adjustment for sampling uncertainty, and the closer will k be to z_{1-p_0} .

The interpretations above are valid also for the combined test in the sub sections page 153 and page 155.

Analogously, considering the formulation (71) of the acceptance criterion in terms of the limiting value, p_{lim} , for the estimated fraction nonconforming, it is seen that in order for the test to provide sufficient assurance that the population proportion nonconforming, $p_a(\mu_d, \sigma_d)$, really is smaller than p_0 , the significance level, α , shall be chosen to be larger than 50%, and consequently the limiting sample fraction nonconforming, p_{lim} , is smaller than p_0 to allow for sampling uncertainty. The larger the sample, the closer p_{lim} will be to p_0 .

This is in line with an interpretation of the statistical test in terms of a confidence interval for $p_a(\mu_d, \sigma_d)$. Acceptance by the statistical test of H_0 :

$p_a(\mu_d, \sigma_d) \leq p_0$ at significance level α is equivalent with accepting whenever the upper α confidence limit for $p_a(\mu_d, \sigma_d)$ does not exceed p_0 . Hence, in order for the test to provide assurance that $p_a(\mu_d, \sigma_d) \leq p_0$, the significance level, α , of the test shall be larger than 50%. In the tables provided in PDA Technical Report No. 25 [22], an assurance, $\alpha = 0.90$ has been suggested.

In terms of acceptance sampling, a requirement of a 90% assurance corresponds to specifying the quality, p_0 as the quality that should have a very low probability of acceptance, $P_{acc} = 0.10$, under the acceptance sampling procedure. Such a quality is often termed *Limiting quality*, (LQ_{10}), indexed by the specified low acceptance probability.

5.6 Discussion

Disregarding the requirement to the sample coefficient of variation in the USP 21 content uniformity test, it was found in Section 5.2 that acceptance probability under the three-class attribute requirements on the individual measurement values in the two-stage procedure only depends upon the proportion of values in the population outside the innermost set of limiting values $LC \pm \Delta$ (e.g. $\Delta = 0.165$). In Section 5.2 a table of the relation between this proportion of population values outside $LC \pm \Delta$ and the acceptance probability under the USP 21 test was provided showing that a requirement of a 95% probability of passing the USP test is equivalent with a requirement that no more than 1.65% of the population values are outside the limits $LC \pm \Delta$. Thus, a requirement on the probability of passing the USP-test may be translated to a population requirement to the proportion of values in the population outside the innermost limits for individual values in the USP-test.

The statistical procedures for acceptance sampling by variables have been tailored for assessing such population values, and the assurance provided by these procedures under due consideration to the sampling uncertainty has been discussed in the statistical literature. An operational acceptance criterion may be formulated in terms of a trapezoidal acceptance polygon in (\bar{d}, s_d) -space that approximates the criterion (71) on the estimated proportion of values outside the limits.

The acceptance polygon criterion may be formulated in various equivalent ways. In particular the criterion may be phrased in terms of a requirement to an *acceptance value* (78), i.e. linear combination of sample average and standard deviation combined with a requirement to the maximal value of the sample standard deviation.

6 Assessment of the properties of the USP pre-view dosage uniformity test

As part of the effort for global harmonization of Uniformity of Dosage Units tests, various changes to the USP general test chapter *Uniformity of Dosage Units* < 905 > have been suggested. The first USP suggestion of changes [27] were modelled after the test provided in the Japanese Pharmacopeia that was based upon such procedures for acceptance sampling by variables that are used in the International Standard ISO 3951 [35] for acceptance sampling by variables for proportion nonconforming items (equivalent to [56]).

Responding to this suggestion, the Statistics Working Group of the Pharmaceutical Research and Manufacturers of America (PhRMA) proposed an alternative that provided the Japanese, the European and the US Pharmacopeia with a harmonized Content Uniformity Test [29], [62]. During the review process various amendments have been introduced resulting in the USP draft proposal [28] published in 2001 as a stage 4 draft.

In the following the statistical properties of the acceptance criteria for uniformity of dosage units by the *Content Uniformity method* will be investigated.

We shall consider the acceptance criteria only for such situations where the target test sample amount at the time of manufacture equals the label claim (LC = 100%, or, in fraction LC = 1).

The proposal has been formulated in terms of an “acceptance value”, and limits for individual units in the sample as follows:

“Calculate the acceptance value as

$$A = A(\bar{d}, s_d) = |M - \bar{d}| + ks_d \quad (80)$$

with

$$k = k(n) = \begin{cases} 2.4 & \text{when } n = 10 \\ 2.0 & \text{when } n = 30 \end{cases} \quad (81)$$

and

$$M = M(\bar{d}) = \begin{cases} \max\{\bar{d}, 0.985\} & \text{when } \bar{d} \leq 1.0 \\ \min\{\bar{d}, 1.015\} & \text{when } 1.0 < \bar{d}. \end{cases} \quad (82)$$

The requirements are met if the acceptance value of the first 10 dosage units is less than or equal to 0.15 (denoted L_1). If the acceptance value is greater than 0.15, test the next 20 units and calculate the acceptance value.

The requirements are met if the final acceptance value of the 30 dosage units is less than or equal to 0.15 (L_1), and no unit is over the deviation of 0.25 (denoted L_2) from the calculated value of $M(\bar{d})$.”

Two different statistical issues are involved in these requirements

- a) In each stage, the criteria include as well a parametric test based upon the acceptance value, (inspection by variables), as limits on individual measurements (inspection by attributes)².

As the requirements in the attribute test involves an “accept zero” criterion, we shall use the term “non-satisfactory units” for units with measurements outside the limits for individual units.

- b) The plan is a two-stage plan with a decision on whether to invoke the second stage depends on the result in the first stage, and the criteria to be applied in the second stage depends on the results in the combined sample (30 units)

²In [28] the attribute criterion is applied only in the second stage. However, as the attribute criterion also was applied in the first stage in the previous draft [29] and [62] we have chosen this criterion also in the first stage.

As the criteria involve the individual values of the 30 sample units, a direct assessment of the operating properties of the test by analytical methods requires numerical integration over a region in a 30-dimensional space, and therefore we have chosen to assess the properties using numerical simulation.

However, before reporting the results of the simulations, the criterion on the acceptance value will be further discussed.

6.1 The acceptance value

Rearranging (82) one finds

$$M(\bar{d}) = \begin{cases} 0.985 & \text{when } \bar{d} < 0.985 \\ \bar{d} & \text{when } 0.985 \leq \bar{d} \leq 1.015 \\ 1.015 & \text{when } 1.015 < \bar{d} \end{cases} \quad (83)$$

and, hence the criterion

$$A(\bar{d}, s_d) \leq L_1 \quad (84)$$

on the calculated acceptance value is equivalent with the set of criteria

$$\begin{aligned} \text{when } \bar{d} < 0.985 & : (0.985 - \bar{d}) + ks_d \leq L_1 \\ \text{when } 0.985 \leq \bar{d} \leq 1.015 & : ks_d \leq L_1 \\ \text{when } 1.015 < \bar{d} & : (1.015 - \bar{d}) + ks_d \leq L_1 \end{aligned} \quad (85)$$

where the inequality to be satisfied in order to pass the criterion (84) depends on the value of \bar{d} .

As the sample standard deviation, $s_d \geq 0$, no value of \bar{d} less than $0.985 - L_1$, or greater than $1.015 + L_1$ will satisfy the criterion, and hence the criterion (84) on the acceptance value may be reformulated as

$$\begin{aligned} \text{when } \bar{d} < 0.985 - L_1 & : \text{not applicable} \\ \text{when } 0.985 - L_1 \leq \bar{d} < 0.985 & : (0.985 - \bar{d}) + ks_d \leq L_1 \\ \text{when } 0.985 \leq \bar{d} \leq 1.015 & : ks_d \leq L_1 \\ \text{when } 1.015 < \bar{d} \leq 1.015 + L_1 & : (1.015 - \bar{d}) + ks_d \leq L_1 \\ \text{when } 1.015 + L_1 < \bar{d} & : \text{not applicable.} \end{aligned} \quad (86)$$

Thus, the criterion on the acceptance value determines a trapezoidal region in (\bar{d}, s_d) -space with corners in the points

$$\begin{aligned}
 A : & \quad \bar{d} = 0.985 - L_1 ; \quad s_d = 0 \\
 B : & \quad \bar{d} = 0.985 ; \quad s_d = L_1/k \\
 C : & \quad \bar{d} = 1.015 ; \quad s_d = L_1/k \\
 D : & \quad \bar{d} = 1.015 + L_1 ; \quad s_d = 0.
 \end{aligned} \tag{87}$$

Comparing the region defined by (87) with the trapezoidal acceptance region derived in the subsections page 153 and page 155, it is seen that letting $\Delta = L_1 + 0.015$ and $k(p_0)$ be given by (81), the two approaches lead to trapezoids with coinciding oblique sides. They only differ in the position of the upper horizontal line specifying the maximum sample standard deviation. Hence, on each stage, the criterion on the acceptance value essentially controls the proportion of population units outside the interval $LC \pm (0.015 + L_1)$

Table 3 summarizes the criteria in terms of sample average and standard deviation (\bar{d}, s_d) , and limits for individual measurements³.

Thus, the procedures in the proposal represent a shift of paradigm compared to the procedures in the 1984 USP 21 [42]. Most notably, the role of the sample average has changed from serving only a normalizing quantity for the sample standard deviation (to form the sample coefficient of variation) to serve now a descriptor of the location of the distribution used in the acceptance value to monitor the proportion of units outside the interval $LC \pm 0.165$ more effectively than the previous attribute criterion on values in the “amber” zone.

6.2 The simulation study

Overall probability of acceptance

Under the distributional assumption (5), the distribution of individual sample values depends only on population mean and standard deviation, (μ_d, σ_d) .

³In the simulations, we have used the sample average in the first sample to determine the limits to be applied to measurements in the first sample, and (when applicable) the total sample average of all 30 measurements to be applied to the subsequent 20 measurements.

Value of \bar{d}	Acceptance limits for s_d	Limits for individual units in the sample
Stage 1, 10 units		
$\bar{d} < 0.835$	-	-
$0.835 \leq \bar{d} \leq 0.985$	$s_d \leq (\bar{d} - 0.835)/2.4$	0.74 - 1.23
$0.985 \leq \bar{d} \leq 1.015$	$s_d \leq 0.0625$	$0.75\bar{d} - 1.25\bar{d}^\dagger$
$1.015 \leq \bar{d} \leq 1.165$	$s_d \leq (1.165 - \bar{d})/2.4$	0.76 - 1.27
$1.165 < \bar{d}$	-	-
Stage 2, all 30 units		
$\bar{d} < 0.835$	-	-
$0.835 \leq \bar{d} \leq 0.985$	$s_d \leq (\bar{d} - 0.835)/2.0$	0.74 - 1.23
$0.985 \leq \bar{d} \leq 1.015$	$s_d \leq 0.075$	$0.75\bar{d} - 1.25\bar{d}$
$1.015 \leq \bar{d} \leq 1.165$	$s_d \leq (1.165 - \bar{d})/2.0$	0.76 - 1.27
$1.165 < \bar{d}$	-	-

Table 3: Acceptance limits for sample standard deviation s_d and limits for individual sample units at each stage in the dosage uniformity test

Therefore the operating properties of the criteria for a distribution of population values characterized by its mean and standard deviation, (μ_d, σ_d) , may be assessed by simulating random samples from a normal distribution with mean and standard deviation (μ_d, σ_d) .

The simulations were performed in a lattice in (μ_d, σ_d) -space, with 3000 samples (each consisting of 30 units) simulated in each lattice point. Hence, the uncertainty due to simulation is at most ± 0.02 (for values of the acceptance probability in the neighbourhood of $P_{acc} = 0.5$) and less than ± 0.0001 when $P_{acc} \geq 0.75$, or $P_{acc} \leq 0.25$.

Figure 8 shows the level-curves (in (μ_d, σ_d) -space) for the overall probability of acceptance under the procedure. The shape of the curves resembles the curves of the solution to $p_{nonc}(\mu_d, \sigma_d; \Delta) = p$ (see (40) and Figure 6) for $\Delta = 0.165$. For comparison, Figure 9 shows the solution to $p_{nonc}(\mu_d, \sigma_d; 0.165) = p$ for various values of p . Superimposing the graphs in the two figures it is seen that there is a good agreement between the shapes of the two set of curves. Thus, the criterion on the acceptance value that at each stage controls the proportion of population units outside the interval $LC \pm (0.015 + L_1)$, apparently also has a dominant effect on the properties of the overall two-stage procedure with supplementary attribute criteria on individual measurement values, d_i .

Comparing the graphs in the two figures it is seen that populations with a proportion $p_{nonc}(\mu_d, \sigma_d; 0.165) = 0.005$ of values outside the interval 1.0 ± 0.165 will have a probability of more than 99% of being accepted by the procedure; when the population proportion of values outside the interval 1.0 ± 0.165 is 0.09, the acceptance probability is 10%.

For each lattice point (μ_d, σ_d) , the value of the acceptance probability $P_{acc}(\mu_d, \sigma_d)$ has been plotted against the corresponding value of $p_{nonc}(\mu_d, \sigma_d; 0.165)$ in Figure 10. The rather narrow sigmoid-shaped scatter of points confirms the impression from Figure 8 that the criteria essentially controls the population proportion of values outside the interval 1.0 ± 0.165 .

The graph allows for an assessment of the approximate probability of acceptance corresponding to any given value of the population proportion of values

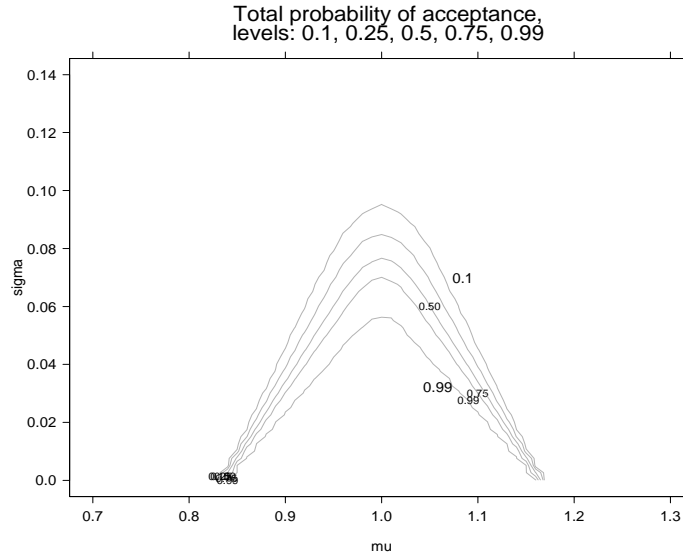


Fig. 8: Level-curves for the overall probability of acceptance for the procedure.

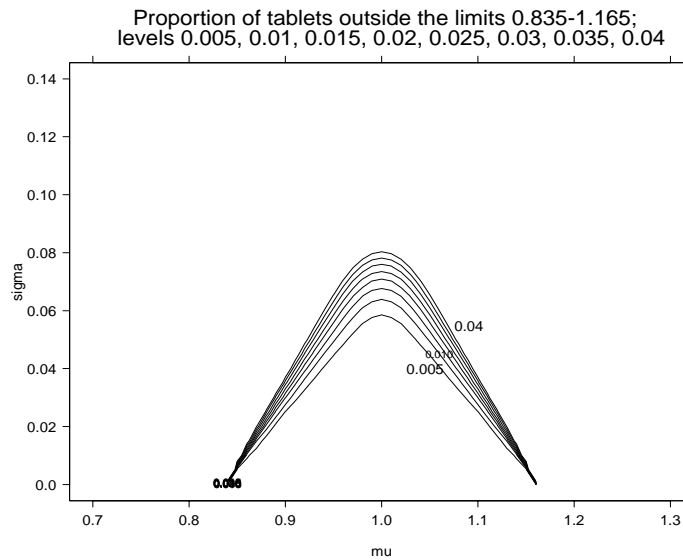


Fig. 9: Combinations, (μ_d, σ_d) of population mean and standard deviation corresponding to specified values, $p_{nonc}(\mu_d, \sigma_d; 0.165)$ of the proportion of population values outside the interval 1.0 ± 0.165 .

outside the interval 1.0 ± 0.165 as indicated in the table below.

p_{nonc}	0.006	0.010	0.075	0.10
$P_{acc}(p)$	0.95	0.90	0.10	0.05

Thus, whenever a batch is accepted by the procedure, there is a confidence of 90% that no more 7.5% of the units in the batch have measurement values outside the interval $LC \pm 0.165$.

However, as the scatter in the vertical direction is larger than what can be explained by simulation uncertainty, there is not a unique value of the acceptance probability corresponding to a given value of $p_{nonc}(\mu_d, \sigma_d)$, but rather an *interval* of values with the specific value of the acceptance probability depending on the particular combination (μ_d, σ_d) giving rise to this proportion nonconforming. Thus, the acceptance probabilities are represented by an OC “band” rather than a single OC-curve. It is well-known that even in the case of single sampling plans such an OC band is an inherent feature of the statistical test, but the thickness of the band depends on the test method selected (see Lei and Vardeman [60]). The thickness of the band clearly has an effect on the steepness of the OC-curve. The thicker the band the less steep is the OC-curve.

For the two-stage plan under study it is believed that the thickness is further enhanced by the choice of the position of the upper horizontal line of the trapezoidal acceptance region in the two stages. Moreover, as discussed in the subsection below (page 167), the additional attribute-criterion on individual measurement values further contribute to the thickness of the band.

The discriminatory effect of the limits for individual measurements

Grossly speaking, the limits to be applied for the individual measurement values correspond to an interval $\pm 25\%$ (termed L_2) around the label claim.

In order to obtain an overall impression of the protection against population proportion of nonsatisfactory units (i.e. with values outside the interval $LC \pm$

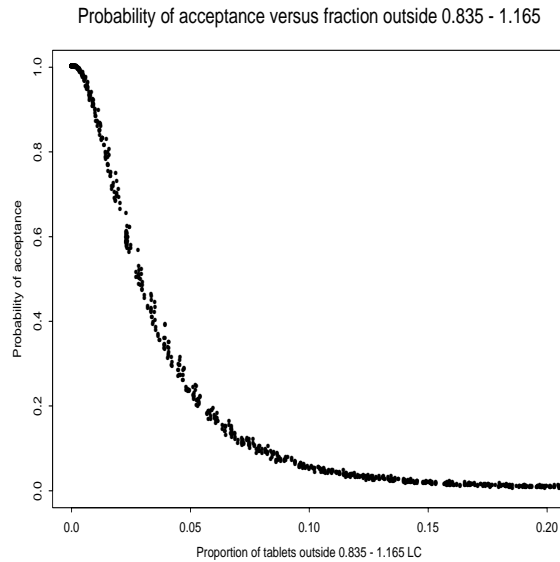


Fig. 10: Overall probability of acceptance vs. the proportion, $p_{nonc}(\mu_d, \sigma_d; 0.165)$, of population values outside the interval 1.0 ± 0.165 .

0.25), Figure 11 shows the solution to $p_{nonc}(\mu_d, \sigma_d; 0.25) = p$ for various values of p . Superimposing the graphs in Figure 11 on the graph of the level-curves of the OC-surface of the procedure (Figure 8) it is seen that the discriminatory power against proportion of nonsatisfactory units depends strongly upon the particular combination of population mean and standard deviation, (μ_d, σ_d) . In some cases, populations with 0.01 % nonsatisfactory units will be rejected with a high probability (larger than 0.9) whereas in other cases (depending on the value of μ_d) such populations will be accepted with a probability higher than 0.99.

However, in greater detail as seen from Table 3 (page 164), the limits depend on the value of the sample average, thereby extending the range of values to the slightly asymmetric interval $0.74 - 1.27$. Moreover, as the limit to be applied depend on the sample average \bar{d} that is subject to random error, the criteria on the individual limits might contribute some random “noise” to the acceptance value criterion and result in a less steep OC-surface.

In order to assess the effect of the attribute criteria on the discriminatory power

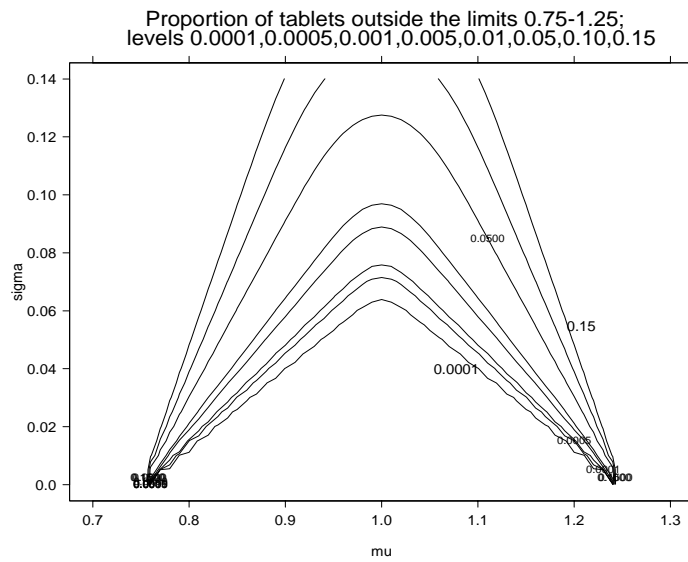


Fig. 11: Combinations, (μ_d, σ_d) of population mean and standard deviation corresponding to specified values, $p_{nonc}(\mu_d, \sigma_d; 0.25)$ of the proportion of nonsatisfactory units in the population (i.e. with values outside the interval 1.0 ± 0.25).

of the procedure, the simulation results were recalculated disregarding the criteria on individual measurement values and using only the criterion based upon the acceptance value for determining acceptability. Figure 12 shows the level-curves (in (μ_d, σ_d) -space) for the probability of acceptance when disregarding the criteria for the individual measurement values. Superimposing these graphs upon the level-curves in Figure 8, it is seen that there is virtually no difference between the two sets of level-curves. Thus, under the assumption of normally distributed measurement values, the criteria for the individual measurement values does not affect the ability to discriminate between different combinations (μ_d, σ_d) of population values. This is in line with the fact that (\bar{d}, s_d) are jointly sufficient for (μ_d, σ_d) and therefore knowledge about the individual measurement values does not add to the information provided by (\bar{d}, s_d) .

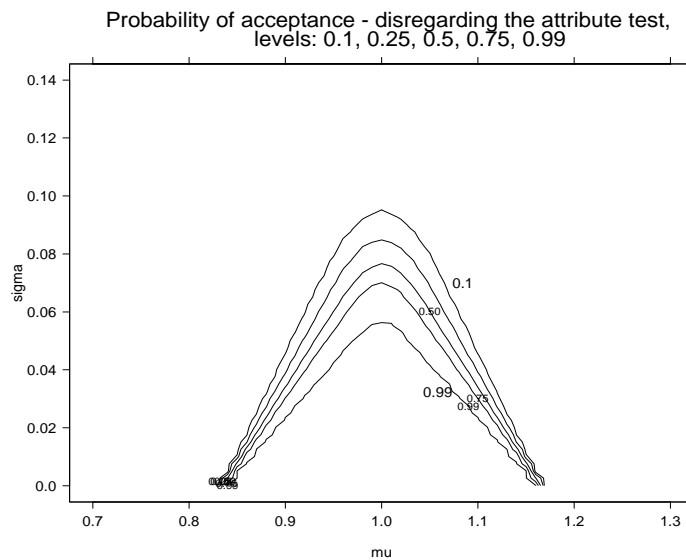


Fig. 12: Level-curves for the overall probability of acceptance for the procedure, disregarding the criteria on individual measurement values.

The two stages

An overview of the procedure and the possible conclusions at each stage has been provided in Figure 13.

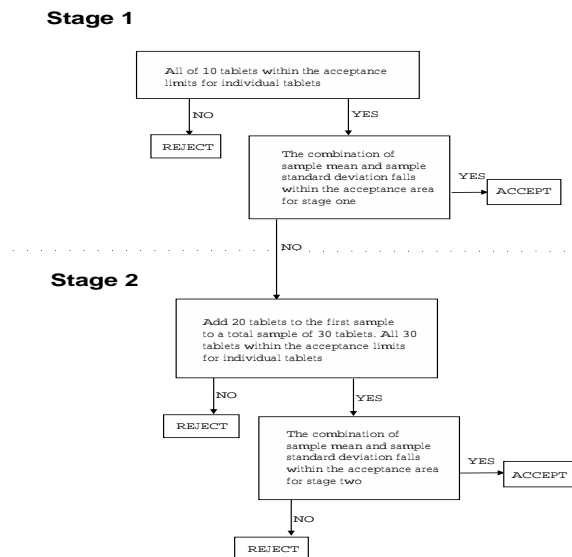


Fig. 13: Schematic representation of the procedure for content uniformity testing.

After assaying the first sample, three different actions are possible

1. **Accept without further testing** when
 - a) all 10 units are within the attribute limits, and
 - b) the calculated acceptance value does not exceed 0.15 (i.e. (\bar{d}, s_d) is within the trapezoidal region)
2. **Test the next 20 units** when
 - a) all 10 units are within the attribute limits, but
 - b) the calculated acceptance value exceeds 0.15 (i.e. (\bar{d}, s_d) is outside the trapezoidal region)
3. **Reject** (i.e. the test is not passed) when one or more units is beyond the attribute limits, irrespective of the acceptance value⁴.

⁴The wording in [28] does not explicitly specify this option; however in previous versions explicit provisions were given for this option. Moreover, the option makes sense in practice, as

Thus, it is possible to conclude the test after testing only 10 units, by outright acceptance (when the acceptance value is satisfactorily small), or by outright rejection (when one or more nonsatisfactory units are found in the first sample).

Figure 14 shows level curves in (μ_d, σ_d) -space of the probability of invoking the second stage of the procedure. Corresponding to a given level of the probability of invoking the second stage there are two curves, viz. one (innermost) curve corresponding to a constant value of $p_{nonc}(\mu_d, \sigma_d; 0.165)$ and another (outermost) corresponding to a constant value of the proportion of nonsatisfactory units, $p_{nonc}(\mu_d, \sigma_d; 0.25)$. This reflects the trade-off between the effect of the criterion on the acceptance value monitoring $p_{nonc}(\mu_d, \sigma_d; 0.165)$, and the attribute criterion $p_{nonc}(\mu_d, \sigma_d; 0.165)$ monitoring $p_{nonc}(\mu_d, \sigma_d; 0.25)$.

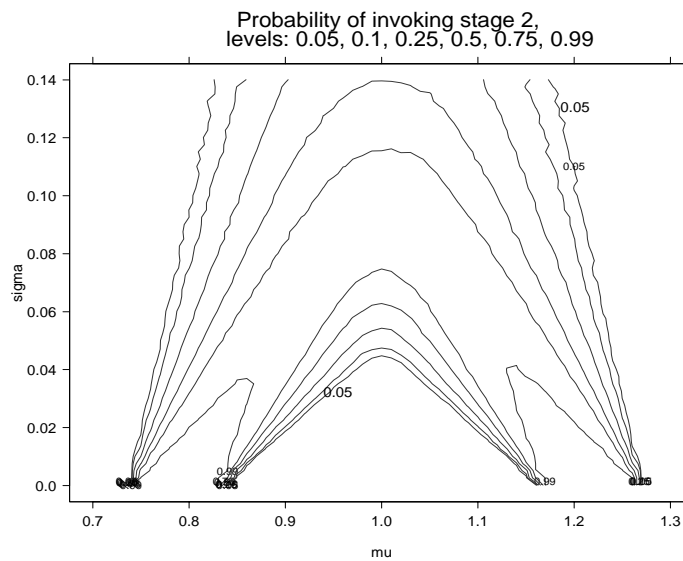


Fig. 14: Level-curves of the probability of invoking the second stage of the procedure.

Populations corresponding to (μ_d, σ_d) -combinations in the inner triangular area with a low probability of invoking the second stage have a high probability

a situation with at least one nonsatisfactory unit found in the first sample would lead to rejection after the second sample, anyhow. Here we have disregarded the extra complication arising from the fact that the attribute limits depend on the sample average, and therefore may change from the first sample to the combined sample.

of being accepted in stage 1, whereas populations corresponding to (μ_d, σ_d) -combinations outside the plotted level-curves with a low probability of invoking the second stage have a high probability of being rejected on stage 1. Populations corresponding to (μ_d, σ_d) -combinations in the “sausage-shaped” area in the middle have a high probability of invoking stage 2, requiring assay of further 20 units.

Comparing with the level-curves of the overall acceptance probability in Figure 8, the effect of giving “a second chance” for such “mediocre populations”, (i.e. populations corresponding to (μ_d, σ_d) -combinations in the middle “sausage-shaped” area that are not accepted in stage 1) may be assessed. Consider e.g. populations corresponding to the innermost 75%-level curve on Figure 14. The curve is seen to correspond to the 50%-level curve of the overall acceptance probability in Figure 8. Thus, such populations are accepted in stage 1 with a probability of 25%, and with 75% probability they are given a second chance, but only one third of these second chances lead to final acceptance.

Thus, from a purely statistical point of view, the test by attributes in stage 1 serves the purpose of saving testing resources for batches that would have been rejected anyhow (after testing all 30 units). This might, however, have been achieved by using also a criterion based upon sample average and standard deviation for rejection in the first stage. Schilling [12] describes the design of two-stage sampling plans by variables that allow for rejection also at stage 1.

As already noted, under the assumption of a normal distribution and independent samples, requirements on individual measurement results are redundant. However, in contemporary industrial applications of acceptance sampling procedures a so-called “accept zero” principle is sometimes considered, viz. a lot can only pass if no nonconforming items are found in the sample, see e.g. [36]. Although the discriminatory power achieved using only such a criterion by attributes is inferior to the power when using sample average and standard deviation, the psychological advantage of invoking this principle is that it conveys a signal that items outside specification are of great concern, and such items should not be found, neither in a sample, nor in the batch. Moreover, the use of an accept zero criterion on individual measurements to supplement

the criterion on sample average and standard deviation has the statistical advantage to make the test procedure more robust towards deviations from the assumption of a normal distribution of dosage content in the tablets.

6.3 Robustness against deviation from distributional assumptions

In order to assess the robustness of the procedure to the assumption of a normal distribution of sample values the simulations were performed also for a lognormal distribution of individual measurements.

The simulations were performed in the same lattice in (μ_d, σ_d) -space as in the previous section. In each lattice point, (μ_d, σ_d) , 3000 samples (each consisting of 30 units) were simulated as random values from a lognormal distribution with that mean and standard deviation.

Figure 15 shows the level-curves (in (μ_d, σ_d) -space) for the overall probability of acceptance under this distributional assumption. Comparing with the level-curves in Figure 8 it is seen that the part of the curves corresponding to values $1.0 < \mu_d$ and larger values of σ_d have been shifted slightly to the right under the lognormal distribution, indicating a slightly larger probability of acceptance under the lognormal distribution than under a normal distribution with the same mean and variance. This may be interpreted by the fact that under the lognormal distribution there is a heavier concentration of probability mass to the left of the mean than under the corresponding normal distribution, and this outweighs the heavier right-hand tail of the lognormal distribution.

However, the heavier tail of the lognormal distribution might imply that the attribute criterion on individual measurement values would have a greater influence than under the normal distribution. In analogy with Figure 12 for normally distributed measurement values, Figure 16 shows the level-curves for the probability of acceptance when disregarding the criteria for the individual measurement values. Superimposing these graphs upon the graph in Figure 15 it is seen that there is virtually no difference between the two sets of level curves. Thus, also under a lognormal distribution of individual measurement

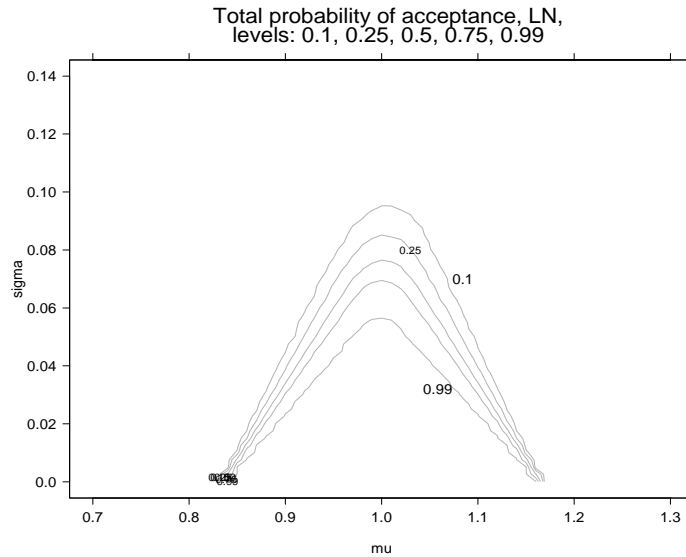


Fig. 15: Level-curves for the overall probability of acceptance for the procedure under a lognormal distribution of measurement values.

values, the criteria on the acceptance values are the dominant criteria of the procedure.

6.4 Equivalent single sampling plan

As an illustration of the direct approach in Section 5.5, consider the determination of a single sampling plan with the same discriminatory properties as the USP preview dosage uniformity test.

The following approximate values for the probability of acceptance as function of the proportion, p , of units outside the interval $LC \pm 0.165$ are read off from Figure 10:

p	0.006	0.010	0.075	0.10
$P_{acc}(p)$	0.95	0.90	0.10	0.05

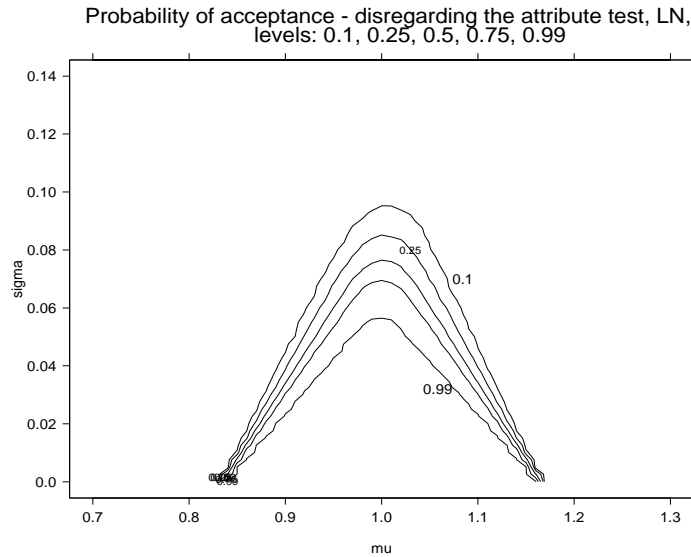


Fig. 16: Level-curves for the overall probability of acceptance for the procedure under a lognormal distribution, disregarding the criteria on individual measurement values

When a single point, $(p, P_{acc}(p))$ on the OC-curve has been specified, it follows from the considerations in Section 5.5, that such a specification defines a relation between sample size n and limiting value, $p_{lim} = p_{lim}(n)$, of the estimated proportion nonconforming units such that the OC-curves of all such sampling plans with (n, p_{lim}) satisfying this relation will pass through the specified point $(p, P_{acc}(p))$. The larger the sample size, the steeper the OC-curve.

Correspondingly, when two points $(p_a, 1 - \alpha)$ and (p_r, β) on the OC-curve have been specified, it is possible to determine a unique combination, (n, p_{lim}) such that the OC-curve for that sampling plan satisfies

$$P_{acc}(p_a) \geq 1 - \alpha, \text{ and } P_{acc}(p_r) \leq \beta$$

i.e. that the plan provides at least the protection specified.

Often p_a is termed the “producer’s risk quality” with α denoting the corresponding producer’s risk, and p_r is termed the “consumer’s risk quality” with

β denoting the corresponding consumer's risk.

In terms of assurance, $1 - \beta$ denotes the assurance that an accepted batch will have a proportion nonconforming that does not exceed p_r .

The table below shows the single sampling plans “matching” the acceptance procedures in the USP preview dosage uniformity test for various choices of matching points, $(p_a, 1 - \alpha)$ and (p_r, β) . It is seen that the discriminatory power of the 10 – 20 two-stage plan corresponds to a single sampling with sample size slightly larger than 20 units. Thus, for good quality productions the savings when using the two-stage plan is 50% (corresponding to acceptance in stage 1), and the extra effort when analysis of the second sample is called for, also amounts to 50% of the sample size for the single sampling plan.

p_a	α	p_r	β	n	k	s_{max}	p_{lim}
0.010	0.10	0.10	0.05	22	1.89	0.076	0.0299
0.006	0.05	0.10	0.05	21	1.91	0.075	0.0278
0.010	0.10	0.075	0.10	24	1.90	0.075	0.0278

The OC-curves have been depicted in Figures 17 to 19.

7 Further issues

We have only discussed situations with independent, identically distributed measurement values where (μ_d, σ_d) may be considered to reflect the population mean dose and the dispersion of doses in the population. This assumption represents an “ideal” situation that does not necessarily reflect all situations occurring in practice. In practice, some other factors might influence the assay mean and the dispersion of the population of potential assay values.

One such factor could be the variation introduced by the analytical procedure. The effect of measurement error on the operating properties of procedures for acceptance sampling by variables has been discussed in [31] and [32].

acceptance probability for plim (lower curve), and trap. region(upper curve)

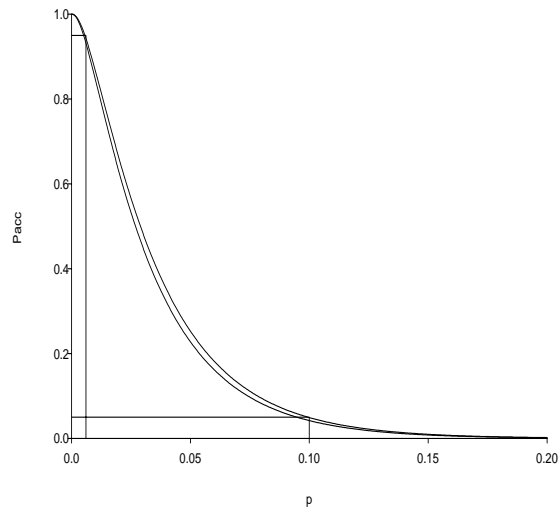


Fig. 17: OC-curve for single sampling plan with $n = 21$, $k = 1.91$, $s_{max} = 0.075$; $p_{lim} = 0.0278$.

acceptance probability for plim (lower curve) and trap. region (upper curve)

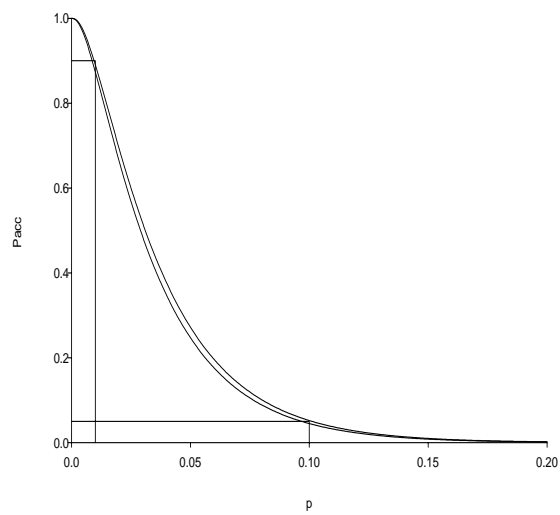


Fig. 18: OC-curve for single sampling plan with $n = 22$, $k = 1.89$, $s_{max} = 0.076$; $p_{lim} = 0.0299$.

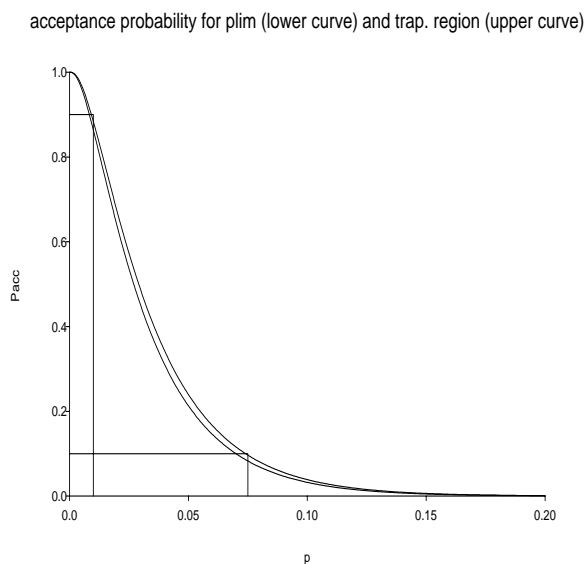


Fig. 19: OC-curve for single sampling plan with $n = 24$, $k = 1.90$, $s_{max} = 0.075$; $p_{lim} = 0.0278$.

Another important factor could be the sampling design used. In practice, in particular in blend sampling, a so-called nested design is sometimes used, where e.g. $n = 36$ sample values are obtained by repeatedly sampling subdividing from only 12 locations. [22] and [62] presents examples showing the operating characteristics of the procedures in the draft USP-proposal under various assumptions on the magnitude of the within location and between location variation. A general discussion of the operating characteristics of a single sampling plan under a nested sampling scheme has been given in [63].

8 Discussion

In the paper we have discussed the statistical properties of various testprocedures used for content uniformity and blend uniformity analysis.

Traditionally, in pharmaceutical regulatory practice such procedures have specified limiting values of sample statistics rather than requirements to population

values. Thus, considerations on the sampling uncertainty have mainly been implicitly considered in the specification of the limiting values, and the assurance provided by the acceptance criteria has not been very transparent.

To overcome this deficiency, various approaches using acceptance criteria based upon prediction intervals or statistical tolerance intervals for future samples have been suggested in the pharmaceutical literature.

In the paper we have discussed the acceptance criteria in terms of the statistical hypothesis concerning population values that is implicitly underlying the criteria. Using the concepts from statistical theory of hypothesis testing, a more transparent description of the assurance provided by the criteria, and the considerations on the sampling uncertainty is obtained.

In particular, we have studied the statistical properties of the acceptance criteria for uniformity of dosage units in the USP draft proposal [28]. The criteria have been related to procedures for acceptance sampling by variables as described in the statistical literature and the individual components of the procedures have been interpreted in terms of concepts from theories of acceptance sampling. The overall properties of the USP draft proposal have been assessed by means of simulation, and as an example of the use of theories and procedures from theories of acceptance sampling a single sampling plan with operating characteristics matching this proposal has been determined.

9 List of symbols

Symbol	
$E[X]$	Mean of the population distribution of X
$V[X]$	Variance of the population distribution of X
D	Relative dose i.e. mass of drug (in blend sample, or in tablet) as fraction, or percentage of target value (random variable)
d	Relative dose in sample unit (actual value)
μ_d	Mean relative dose (population value)
σ_d	Standard deviation in distribution of relative doses (population value)
C_d	Coefficient of variation in distribution of relative doses (population value) $C_d = \sigma_d / \mu_d$
n	Number of units in sample
\bar{D}	sample average relative dose per sample unit (random variable)
\bar{d}	sample average relative dose per sample unit (actual value)
S_d	sample standard deviation of relative doses in sample (random variable)
s_d	sample standard deviation of doses in sample (actual value)
LC	Required dose, Label Claim. In Section 5 and 6 measurements are relatively to LC, i.e. LC=1 (100%).
χ_f^2	Random variable distributed according to a χ^2 -distribution with f degrees of freedom
χ_f, p^2	p 'th quantile in χ^2 -distribution with f degrees of freedom, (i.e with probability mass p to the left of this value) Some (mainly US) textbooks use the notation $\chi_{\alpha, \nu}^2$ to denote the so-called χ -squared critical value, denoting the number on the measurement axis such that the probability mass for the χ_{ν}^2 -distribution to the right of this value is α .
$F(f_1, f_2)$	Random variable distributed according to a F-distribution with (f_1, f_2) degrees of freedom
$F(f_1, f_2)_p$	p 'th quantile in F-distribution with (f_1, f_2) degrees of freedom, (i.e with probability mass p to the left of this value)

Symbol	
$t_f(\delta)$	Random variable distributed according to a noncentral t-distribution with f degrees of freedom and non-centrality parameter δ
$t(f, \delta)_p$	p 'th quantile in noncentral t-distribution with f degrees of freedom and non-centrality parameter δ , (i.e with probability mass p to the left of this value)
z_p	p 'th quantile in standard normal distribution, (i.e with probability mass p to the left of this value)
Δ, Δ_1	quantity serving to specify the limits for individual values (usually in the form of $LC \pm \Delta$)
$P_{3-c}(\mu_d, \sigma_d)$	Acceptance probability under a 3-class attribute one-stage sampling plan when mean and standard deviation in the population are (μ_d, σ_d)
$p_{nonc}(\mu, \sigma; \Delta)$	Generic function (34) expressing the probability mass outside the limits $LC \pm \Delta$ in a normal distribution with mean μ and standard deviation σ .
$P_{usp}(\mu_d, \sigma_d)$	probability of passing the USP-21 test when mean and standard deviation in the batch is (μ_d, σ_d)
$p_a(\mu_d, \sigma_d)$	fraction of units outside limits $LC \pm \Delta$ when mean and standard deviation in the batch is (μ_d, σ_d)
$p^+(\mu_d, \sigma_d)$	fraction of units violating upper limit, $LC + \Delta$ when mean and standard deviation in the batch are (μ_d, σ_d)
$p(\mu_d, \sigma_d)$	fraction of units violating lower limit, $LC - \Delta$ when mean and standard deviation in the batch are (μ_d, σ_d)
$A(\bar{d}, s_d)$	Acceptance value used for determining acceptability, see (80)

Paper E

On particle size distributions and dosage uniformity for low-dose tablets

E

1 Introduction and summary

In pharmaceutical production of tablets it is of natural concern that the active ingredient is distributed uniformly among the individual units of the batch. Therefore, during production various tests are performed to ensure uniformity of the blend and uniformity of content in the final product.

The statistical properties of such tests of content uniformity are usually assessed assuming a normal distribution of content in the tablets. However, actual distributions of particle sizes are often seen to be skewed, and it is therefore conceivable that this feature will also affect the shape of the distribution of content in the dosage units. In this paper the particles referred to are the drug particles in the blend.

In the paper we investigate the effect of relative variation, skewness and excess (heavy-tailedness) of the distribution of particle diameters on the resulting distribution of particle volume and particle mass under the assumption that particles are spherical. For a log-normal distribution of particle diameters, the resulting distribution of particle volume and mass is also a log-normal distribution with a mean that is larger than the mass corresponding to the mean radius, and a coefficient of variation that is larger than the coefficient of variation for the distribution of the radii. It is shown that this implies that the skewness and excess of the distribution of particle radii is amplified when transformed to the distribution of particle mass. The larger the coefficient of variation in the distribution of particle radii, the more these departures from normality are amplified.

Along with the variation in particle mass, an important source of dose variation for low-dose tablets is the variation in the number of drug particles in the tablets. We investigate the transfer of the variation of particle sizes and of number of particles in the tablets to dose variation in tablets under the assumption of a homogeneous blend with a random scattering of particles over the blend, and derive expressions for skewness and excess of the distribution of tablet doses. It is demonstrated (as previously shown by Yalkowsky and Bolton [64]) that for a given distribution of particle sizes the variance in the distribution of

absolute doses is proportional to the average number of drug particles in the tablets, and moreover we show that the larger the average number of particles in the tablets, the closer will the distribution of tablet doses be to a normal distribution.

It is a common practice in the pharmaceutical industry to perform tests for content uniformity on measures of dispersion of relative doses (relative to label claim) rather than on absolute doses. Translating the above results to relative content it follows that the standard deviation of the relative content in tablets decreases when the average number of particles in the tablets is increased. Thus, for a given distribution of particle sizes in the blend, the standard deviation of the relative content depends on the label claim (or target value). The smaller the label claim, the larger is the relative standard deviation, and the more pronounced is the departure from a normal distribution of relative doses.

In practical production, the assumptions of an ideal blend (random scattering of drug particles) is not always satisfied, but some clustering of particles is conceivable. We propose a simple and transparent hierarchical model for particle distribution that reflects a varying intensity of particles over the blend. The properties of the distribution of relative doses under this model are derived.

Although the results are presented in terms of content uniformity of tablets, the statistical results are also valid for blend samples, under the assumption that no systematic sampling error is introduced by the sampling process. The only difference being that the size of individual blend samples usually is larger than the size of the dosage units, and therefore in the interpretation of the results, label claim shall be substituted by the target value of the blend sample.

2 Lognormal distribution of particle radii

Assume that the distribution of particle sizes may be adequately described by a log-normal distribution. In pharmaceutical practice this distributional assumption is generally considered to provide a good description of particle size data.

Furthermore, the use of the log-normal distribution to characterize particle size data has the advantage that this assumption allows for explicit mathematical expressions for the distribution of spherical particle volumes.

An implicit assumption underlying the use of the log-normal distribution model is that the sieving/milling process has been satisfactory in the sense that no single, large particles are present.

2.1 Distribution of particle radii

Let R denote the radius of a particle. The assumption of a log-normal distribution of particle radii is equivalent to assuming that the distribution of the logarithm of the radii is a normal distribution.

The probability density function for a log-normal distribution is

$$f_R(r) = \frac{1}{\ln(r) \ln(\beta_R) \sqrt{2\pi}} \exp\left(-\frac{1}{2} \left[\frac{\ln(r) - \alpha_R}{\beta}\right]^2\right) \quad (1)$$

where the parameters α_R and β_R are related to the so-called geometric mean, μ_R^g , and the geometric standard deviation by

$$\mu_R^g = \exp(\alpha_R) \quad (2)$$

$$\sigma_R^g = \exp(\beta_R). \quad (3)$$

For the purpose of interpretation it is usually more convenient to work in terms of the moments of the distribution,

$$\mu_R = \mathbf{E}[R] = \exp(\alpha_R + (1/2)\beta_R^2) \quad (4)$$

$$\sigma_R = \sqrt{\mathbf{V}[R]} = \mu_R \times \sqrt{\exp(\beta_R^2) - 1}. \quad (5)$$

The coefficient of variation is

$$C_R = \sigma_R / \mu_R = \sqrt{\exp(\beta_R^2) - 1}. \quad (6)$$

Hence, the parameter $\beta_R = \ln(\sigma_R^g)$ is in a one-to-one correspondence with the coefficient of variation, C_R

$$\beta_R = \sqrt{\log(1 + C_R^2)} \quad (7)$$

$$\alpha_R = \log \mu_R - \frac{1}{2} \log(1 + C_R^2) \quad (8)$$

and

$$\exp(\beta_R^2) = 1 + C_R^2. \quad (9)$$

Expressed in terms of the mean, μ_R and the coefficient of variation, the moments about the mean are

$$\mu_2(R) = V[R] = \mu_R^2 C_R^2 \quad (10)$$

$$\mu_3(R) = \mathbf{E}[(R - \mu_R)^3] = \mu_R^3 \times C_R^4 \times (C_R^2 + 3) \quad (11)$$

$$\begin{aligned} \mu_4(R) &= \mathbf{E}[(R - \mu_R)^4] \\ &= \mu_R^4 \times C_R^4 \times \left[(1 + C_R^2)^4 + 2(1 + C_R^2)^3 + 3(1 + C_R^2)^2 - 3 \right] \end{aligned} \quad (12)$$

Hence, the coefficients of skewness, γ_1 , and excess, γ_2 are (see e.g. [45])

$$\gamma_1(R) = \mu_3(R)/\sigma_R^3 = C_R \times (C_R^2 + 3) > 0 \quad (13)$$

$$\begin{aligned} \gamma_2(R) &= \mu_4(R)/\sigma_R^4 - 3 \\ &= (1 + C_R^2)^4 + 2(1 + C_R^2)^3 + 3(1 + C_R^2)^2 - 6 \\ &= C_R^2(16 + 15C_R^2 + 6C_R^4 + C_R^6) > 0. \end{aligned} \quad (14)$$

Thus, the distribution is positively skewed, and has heavier tails than the normal distribution (is leptokurtic).

2.2 Distribution of particle mass for spherical particles

Assuming that particles are spherical, the particle volume, V , is $V = (4\pi/3)R^3$ and, hence the moments in the distribution of particle volume may be expressed in terms of the moments in the distribution of particle radii, R

C_R	γ_1	γ_2
0.1	0.30	0.16
0.2	0.61	0.66
0.3	0.93	1.57
0.4	1.26	2.97
0.5	1.63	5.04
0.6	2.02	8.00
0.7	2.44	12.21
0.8	2.91	18.12
0.9	3.43	26.42
1.0	4.00	38.00
1.2	5.33	76.36
1.4	6.94	148.92
1.6	8.90	282.88
1.8	11.23	523.58
2.0	14.00	944.00

Table 1: Skewness, γ_1 , and excess, γ_2 , for log-normal distributions of particle radii for different values of C_R .

$$\mathbf{E}[R^3] = \mu_R^3 \left(C_R^4 (C_R^2 + 3) + 3C_R^2 + 1 \right) \quad (15)$$

$$\mathbf{V}[R^3] = \mathbf{E}[R^6] - (\mathbf{E}[R^3])^2. \quad (16)$$

For a log-normal distribution of particle radii, the distribution for the particle volume, $V = (4\pi/3)R^3$, is also a log-normal distribution with parameters

$$\alpha_V = \ln(4\pi/3) + 3\alpha_R \quad (17)$$

$$\beta_V = 3\beta_R. \quad (18)$$

Assuming that all drug particles have the same mass density, ρ , the distribution of particle mass $M = \rho V$ is also log-normal with parameters

$$\alpha_M = \ln(4\pi/3) + \ln(\rho) + 3\alpha_R \quad (19)$$

$$\beta_M = 3\beta_R. \quad (20)$$

Hence the mean particle mass is

$$\mu_M = \mathbf{E}[M] = \exp(\alpha_M + (1/2)\beta_M^2) = (4\pi/3)\rho\mu_R^3(1 + C_R^2)^3 \quad (21)$$

and the coefficient of variation is

$$C_M = \sqrt{\exp(\beta_M^2) - 1} = \sqrt{(1 + C_R^2)^9 - 1} \quad (22)$$

with the relation

$$1 + C_M^2 = (1 + C_R^2)^9 \quad (23)$$

$$\sigma_M = \mu_M \times \sqrt{(1 + C_R^2)^9 - 1}. \quad (24)$$

Utilizing the fact that the distribution of particle mass is also a log-normal distribution, the third and fourth moment about the mean are found in analogy with (11) and (12)

$$\mu_3(M) = \mathbf{E}[(M - \mu_M)^3] = \mu_M^3 C_M^4 (C_M^2 + 3) \quad (25)$$

$$\begin{aligned} \mu_4(M) &= \mathbf{E}[(M - \mu_M)^4] \\ &= \mu_M^4 C_M^4 \left[(1 + C_M^2)^4 + 2(1 + C_M^2)^3 + 3(1 + C_M^2)^2 - 3 \right] \end{aligned} \quad (26)$$

In analogy with (13) and (14) one obtains the coefficients of skewness and excess for the distribution of particle mass

$$\gamma_1(M) = \mu_3(M)/\sigma_M^3 = C_M \times (C_M^2 + 3) > 0 \quad (27)$$

$$\begin{aligned} \gamma_2(M) &= \mu_4(M)/\sigma_M^4 - 3 \\ &= (1 + C_M^2)^4 + 2(1 + C_M^2)^3 + 3(1 + C_M^2)^2 - 6 \\ &= C_M^2(16 + 15C_M^2 + 6C_M^4 + C_M^6) > 0. \end{aligned} \quad (28)$$

Inserting μ_M and C_M from (21) and (22) the moments of the distribution of particle mass may be expressed in terms of the parameters in the distribution of particle radii as

$$\begin{aligned} \mu_3(M) &= \mathbf{E}[(M - \mu_M)^3] \\ &= [(4\pi/3)\rho\mu_R^3]^3(1 + C_R^2)^9[(1 + C_R^2)^9 - 1]^2[(1 + C_R^2)^9 + 2] \quad (29) \\ \mu_4(M) &= \mathbf{E}[(M - \mu_M)^4] \\ &= [(4\pi/3)\rho\mu_R^3]^4[(1 + C_R^2)^9 - 1]^2(1 + C_R^2)^{12}[(1 + C_R^2)^9 - 1]^2 \times \\ &\quad [(1 + C_R^2)^{36} + 2(1 + C_R^2)^{27} + 3(1 + C_R^2)^{18} - 3]. \end{aligned} \quad (30)$$

Further, the coefficients of skewness and excess for particle mass may be expressed in terms of the coefficient of variation, C_R in the distribution of the radii as

$$\gamma_1(M) = \left(\sqrt{(1 + C_R^2)^9 - 1} \right) \left((1 + C_R^2)^9 + 2 \right) \quad (31)$$

$$\gamma_2(M) = (1 + C_R^2)^{36} + 2(1 + C_R^2)^{27} + 3(1 + C_R^2)^{18} - 6. \quad (32)$$

Table 2.2 shows the coefficient of variation in the distribution of particle mass and the corresponding coefficients of skewness and excess for different values of the coefficient of variation in the distribution of particle radii. It is seen that the relative variation in the distribution of the radii is amplified when considering the distribution of particle mass, resulting in a similar marked amplification of the coefficients of skewness and excess.

C_R	C_M	γ_1	γ_2
0.1	0.31	0.95	1.64
0.2	0.65	2.23	9.95
0.3	1.08	4.52	50.89
0.4	1.67	9.72	356.55
0.5	2.54	24.00	4.07×10^3
0.6	3.86	69.20	7.30×10^4
0.7	5.93	226.6	1.82×10^6
0.8	9.21	808.8	5.55×10^7
0.9	14.40	3.03×10^3	1.91×10^9
1.0	22.61	1.16×10^4	6.90×10^{10}
1.2	55.36	1.70×10^5	8.84×10^{13}
1.4	132.07	2.30×10^6	9.26×10^{16}
1.6	303.06	2.78×10^7	7.12×10^{19}
1.8	665.50	2.95×10^8	3.85×10^{22}
2.0	1 397.54	2.73×10^9	1.46×10^{25}

Table 2: Coefficient of variation, C_M , skewness, γ_1 , and excess, γ_2 , in log-normal distributions of particle mass for different values of coefficient of variation, C_R , in log normal distribution of particle radii.

3 Distribution of dose content under random mixing of particles

3.1 Modelling random mixing

Most of the probabilistic models for mixtures in the literature discuss statistical properties of samples from mixtures of two (or more) components when sampling a fixed number of particles, or a fixed volume from the mixture, see e.g. the survey by Gjelstrup Kristensen [13] and concern the relative distribution (by mass) of the two components in the sample. In these models the two components enter symmetrically and the main issue is to describe the overdispersion compared to a binomial distribution of particles of the key component.

For low-dose tablets focus is upon the distribution of the small fraction of key component particles in the blend, and this symmetry between key component particles and other particles is not necessarily of concern. Therefore, in the following we shall investigate the distribution of dose content in tablets under various assumptions on the distribution of key component particles in the blend. As the resulting variation in dose content in tablets is caused by variation of the number of drug particles in the sample as well as variation in particle sizes, the discussion aims at separating the contribution from these two sources of variation, thereby allowing for more transparent analytical formulae for the variation.

3.2 Constant number of particles in tablets

The dose content of a tablet is the sum of the masses of the N drug particles in that tablet

$$D = \sum_{i=1}^N M_i \quad (33)$$

with M_i denoting the mass of the i 'th particle.

If all tablets (or samples) contained the same number of drug particles, N , dose

variation would only be caused by the variation in the mass of the individual drug particles selected by the tableting process. Assume that particles are sampled at random from the distribution (by number) of particle radii.

For later reference the moments in the distribution of doses under this simplifying assumption are given below.

$$\mu_{D|N} = \mathbf{E}[\sum_{i=1}^N M_i] = N \times \mu_M \quad (34)$$

$$\mu_2(D|N) = \sigma_{D|N}^2 = \mathbf{V}[\sum_{i=1}^N M_i] = N \times \sigma_M^2 \quad (35)$$

$$\mu_3(D|N) = \mathbf{E}[(D - \mu_{D|N})^3] = \mathbf{E}[(D - N\mu_M)^3] = N\mu_3(M) \quad (36)$$

$$\mu_4(D|N) = \mathbf{E}[(D - \mu_{D|N})^4] = N\mu_4(M) + 3N(N-1)\sigma_M^4 \quad (37)$$

where (37) is found from

$$\kappa_4(D|N) = \mu_4(D|N) - 3(\mu_2(D|N))^2$$

such that

$$\mu_4(D|N) = \kappa_4(D|N) + 3(\mu_2(D|N))^2$$

but

$$\kappa_4(D|N) = N\kappa_4(M) = N(\mu_4(M) - 3\mu_2(M)^2)$$

and

$$\mu_2(D|N) = N\mu_2(M)$$

such that

$$\mu_4(D|N) = N(\mu_4(M) - 3\mu_2(M)^2) + 3N^2\mu_2(M)^2 \quad (38)$$

$$= N\mu_4(M) + 3N(N-1)\mu_2(M)^2. \quad (39)$$

The coefficient of variation in the distribution of doses is found from (34) and (35) as

$$C_{D|N} = \frac{\sigma_{D|N}}{\mu_{D|N}} = \frac{1}{\sqrt{N}} \times C_M \quad (40)$$

and the coefficients of skewness and excess are

$$\gamma_1(D|N) = \frac{1}{\sqrt{N}} \times \gamma_1(M) \quad (41)$$

$$\gamma_2(D|N) = \frac{1}{N} \times \gamma_2(M). \quad (42)$$

For a log-normal distribution of particle masses, $\gamma_1(M)$ and $\gamma_2(M)$ may be expressed in terms of C_M by (27) and (28), respectively.

Spherical particles

For spherical particles with a log-normal distribution of particle radii, the coefficient of variation in the distribution of doses may be expressed in terms of the coefficient of variation, C_R , in the distribution of radii using (22) such that

$$C_{D|N} = \frac{1}{\sqrt{N}} \times \sqrt{(1 + C_R^2)^9 - 1} \quad (43)$$

and $\gamma_1(M)$ and $\gamma_2(M)$ are given by (31) and (32), respectively.

The expressions (41) and (42) show that the effect of skewness and excess of the distribution of particle masses is diminished when particle masses are added to form tablets. The larger the number of particles, the smaller the skewness and excess, and the better will the distribution of doses be approximated by a normal distribution. This is also a consequence of the central limit theorem in probability theory.

3.3 Random variation of number of particles in tablets

In practice, the number of particles in a tablet exhibits variation. For low-dose tablets the drug particles occupy only a small fraction of the blend volume, and therefore under perfect mixing the distribution of dose particles in the three-dimensional space should share the properties of a random distribution of points, i.e. a spatial Bernoulli process, which for practical purposes may be approximated by a spatial Poisson process, see Stoyan et al [65].

When the intensity of drug particles is small, perfect mixing should imply that the distribution of particle volume in a sample unit is independent of the number of particles in the unit. For large particles, however, this assumption is not necessarily satisfied, and more structured random measures need to be invoked, see Stoyan et al [65].

In the present discussion we shall assume that the variation in the number of particles in tablets may be described by independent samples from a Poisson distribution with mean \bar{N} , and that the distribution of particle sizes in a sampled unit correspond to a random sample from the (number) distribution of particle sizes.

Thus, the probability mass function for the number, N , of particles in a tablet is

$$f_N(n) = \frac{(\bar{N})^n}{n!} \exp(-\bar{N}). \quad (44)$$

It is a well-known property of the Poisson distribution (see e.g. Johnson and Kotz, [66]) that the variance, $V[N]$, of the distribution is equal to the mean, $E[N] = \bar{N}$.

For a specified absolute tablet dose, LC, and a drug with mean particle mass, μ_M , the mean number of particles will be (assuming that the mean dose equals label claim, LC)

$$\bar{N} = \text{LC}/\mu_M. \quad (45)$$

Taking the variation of the number N of particles into account, and utilizing that for the Poisson distribution $V[N] = \bar{N}$, the mean and variance of the dose under an arbitrary distribution of particle masses are found as

$$\mu_D = E[D] = E_N[\mu_{D|N}] = E_N[N\mu_M] = \bar{N}\mu_M \quad (46)$$

$$\sigma_D^2 = V[D] = E_N[\sigma_{D|N}^2] + V_N[\mu_{D|N}] = E_N[N\sigma_M^2] + V_N[N\mu_M] \quad (47)$$

$$= \bar{N}\sigma_M^2 + \mu_M^2 V[N] = \bar{N}\mu_M^2 (C_M^2 + 1) \quad (48)$$

and hence the coefficient of variation in the distribution of doses is

$$C_D = \frac{1}{\sqrt{\bar{N}}} \sqrt{1 + C_M^2}. \quad (49)$$

We note that the coefficient of variation, C_D , of dose content under this assumption of a random spatial distribution of particles mixing satisfies

$$C_D = C_{D|\bar{N}} \times \sqrt{1 + 1/C_M^2} \quad (50)$$

with $C_{D|\bar{N}}$ given by (40).

Hence, as in the case of a constant number of particles in all dosage units, the coefficient of variation decreases with the average number of particles as $1/\sqrt{\bar{N}}$. However, the variation of the number of particles in the individual dosage units implies that the coefficient of variation of dose content is larger than the coefficient of variation corresponding to a fixed number of particles in each tablets (for the same average number of particles in tablets under the two schemes). The smaller the coefficient of variation in the distribution of particle mass, the more pronounced is the difference between the two schemes. For large variation of particle masses, the variation of particle masses outweighs the effect of the varying number of particles.

Expressed in terms of label claim (LC) one finds using (45)

$$\sigma_D^2 = V[D] = LC \times \mu_M (C_M^2 + 1) \quad (51)$$

$$C_D = \sqrt{\frac{\mu_M (C_M^2 + 1)}{LC}} \quad (52)$$

which shows that under the assumption of random distribution of particles and a given distribution of particle sizes does the absolute standard deviation increase proportional to the square root of the magnitude of the dose (LC), and hence the coefficient of variation decreases proportional to the square root of the dose. The larger the dose, the smaller a coefficient of variation in the distribution of doses. This is in agreement with practical experience where it is found that problems mainly occur for small dose tablets.

When relative doses, $d = D/LC$, are considered, the variance, σ_d^2 , of the distribution of relative doses is

$$\sigma_d^2 = V[D/LC] = C_D^2 = \frac{\mu_M}{LC} (C_M^2 + 1) \quad (53)$$

with the coefficient of variation, C_d , in the distribution of relative doses given by (52) as $C_d = C_D$.

This is in line with the general observation from theories on blend uniformity that the standard deviation of relative content decreases as the square root of sample weight (or volume).

The relation (53) has previously been established by Yalkowsky and Bolton [64]. The relation is in agreement with the well known results on variance-sample size relations for random binary mixtures. Noting that $\mu_M/\text{LC} = 1/\bar{N}$ the relation (53) simply states that the variance in the distribution of relative doses is proportional to $1/\bar{N}$.

Spherical particles

Under the assumption of spherical particles and a log-normal distribution of particle radii, (21) gives μ_M as a function of mean particle radius, μ_R , and coefficient of variation, C_R , in the distribution of particle radii

$$\mu_M = (4\pi/3)\rho\mu_R^3 \times (1 + C_R^2)^3 \quad (54)$$

which inserted into (45) gives the average number of particles as function of label claim and of the parameters in the distribution of particle radii

$$\bar{N} = \frac{3 \times \text{LC}}{4\pi \times \rho \times \mu_R^3 \times (1 + C_R^2)^3}. \quad (55)$$

For spherical particles, μ_M and C_M are given by (21) and (22),

and one obtains

$$\mu_D = \mathbf{E}[D] = \bar{N}\mu_M = \bar{N} \times (4\pi/3)\rho\mu_R^3 \times (1 + C_R^2)^3 \quad (56)$$

$$\sigma_D^2 = \mathbf{V}[D] = \bar{N} \times \left[(4\pi/3)\rho \right]^2 \times \mu_R^6 (1 + C_R^2)^{15} \quad (57)$$

$$C_D = \frac{\sigma_D}{\mu_D} = \frac{1}{\sqrt{\bar{N}}} \times (1 + C_R^2)^{9/2} \quad (58)$$

or, in terms of label claim, using (45)

$$\sigma_D^2 = \mathbf{V}[D] = \frac{4\pi\text{LC} \times \rho}{3} \times \mu_R^3 (1 + C_R^2)^{12} \quad (59)$$

$$C_D = \sqrt{\frac{4\pi\rho}{3\text{LC}}} \times \mu_R^{3/2} (1 + C_R^2)^6. \quad (60)$$

Coefficient of variation in distribution of doses, LC = 10, rho = 1.3 g/cc

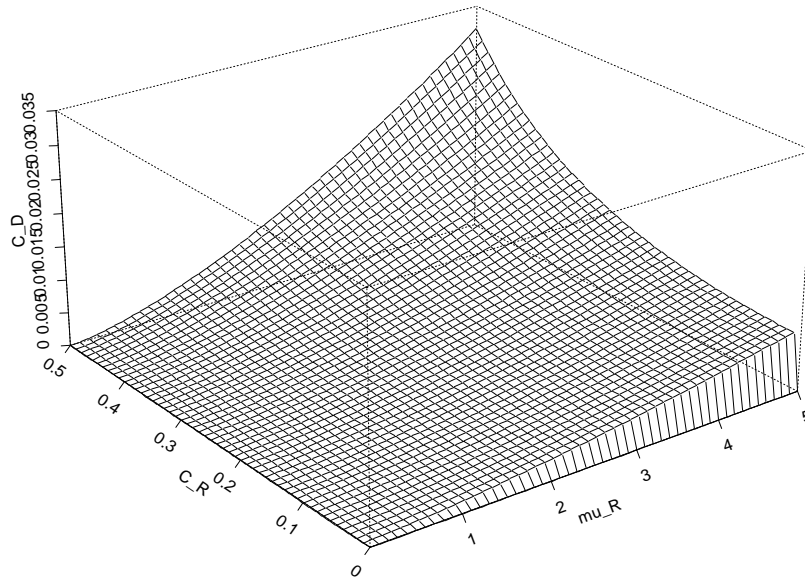


Fig. 1: Coefficient of variation, C_D , in the distribution of doses as function of mean particle radius, μ_R , and coefficient of variation, C_R , in the distribution of particle radii. LC is given in μg and particle radius is given in μm .

When relative doses, $d = D/\text{LC}$ are considered, the variance, σ_d^2 (53) under the assumption of spherical particles may be expressed as

$$\sigma_d^2 = \frac{4\pi \times \rho}{3\text{LC}} \times \mu_R^3 (1 + C_R^2)^{12}. \quad (61)$$

As an illustration, Figure 1 shows the coefficient of variation in the distribution of doses as function of mean particle radius and coefficient of variation in distribution of particle radii when LC = 10 mg and $\rho = 1.3 \text{ g/cm}^3$. It is seen that the coefficient of variation in the distribution of doses increases rapidly with mean particle radius and coefficient of variation in the distribution of radii, reflecting the decrease in average number of drug particles.

3.4 Minimum number of particles necessary to secure a specified dose coefficient of variation

In practice, tests for demonstrating blend uniformity or content uniformity contain requirements on the standard deviation of the content of the samples (blend samples or dosage units) expressed relative to the target value (LC).

It is therefore of interest to assess the implications for the distribution of particle sizes of such requirements to blend or content uniformity. For simplicity only a requirement on the coefficient of variation, C_D will be considered.

For a specified maximum value, C_{max} of C_D , it follows from (52) that under the assumption of perfect mixing, the mean particle mass, μ_M , shall satisfy

$$\mu_M \leq \text{LC} \times \frac{C_{max}^2}{C_M^2 + 1} \quad (62)$$

with C_M denoting the coefficient of variation in the distribution of particle mass.

Thus, the limit for acceptable particle masses increases proportional to LC, and decreases proportional to $(1 + C_M^2)^{-1}$ as C_M increases.

Under the assumption of spherical particles, the limit may be expressed in terms of the coefficient of variation, C_R , for the distribution of particle radii using (60)

$$\mu_R \leq \frac{[C_{max}^2 \times 3\text{LC}/(4\pi\rho)]^{1/3}}{(1 + C_R^2)^4}. \quad (63)$$

Thus, the limit for acceptable particle radii increases proportional to the third root of LC, and decreases with increasing coefficient of variation, C_R , for the distribution of particle radii, the increase being proportional to $(1 + C_R^2)^{-4}$.

Reducing the label claim by 50% implies that the maximum allowable average particle radius is reduced by 20% (if the coefficient of variation in the distribution of particle radii remains unchanged). Reducing the label claim to a tenth implies that the maximum allowable average particle radius is reduced by 50%.

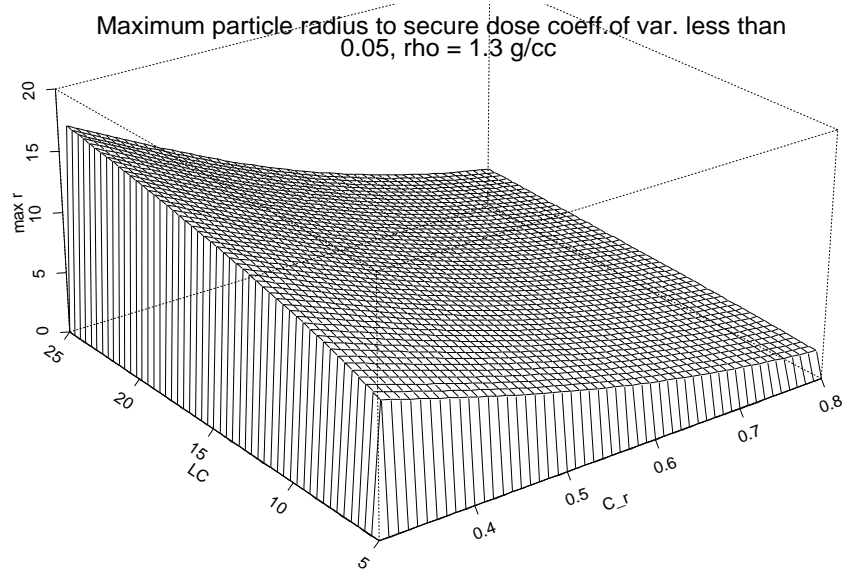


Fig. 2: Maximum mean particle radius necessary to secure a coefficient of variation in the distribution of dosage values less than $C_{max} = 0.05$ shown as function of label claim and coefficient of variation in distribution of particle radii. LC is given in μg and particle radius is given in μm .

Thus, a blend that might produce a satisfactory distribution of doses (in terms of the coefficient of variation in the distribution of relative doses) for large dose tablets need not be satisfactory for smaller dose tablets.

Figure 2 shows the limit (63) for particle radii necessary to assure that the coefficient of variation in the distribution of doses does not exceed $C_{max} = 0.05$. The limit is shown as function of label claim, LC, and coefficient of variation, C_R , in the distribution of particle radii for $\rho = 1.3 \times 10^{-6} \mu g / \mu m^3$. It is seen that the maximal radius increases rather slowly with the label claim (proportional to the third root of label claim).

The requirement (62) to the mean drug particle size may be expressed as a

requirement to the average number of particles in a dosage unit:

$$\bar{N} \geq \frac{(1 + C_M^2)}{C_{max}^2} = \frac{(1 + C_R^2)^9}{C_{max}^2}. \quad (64)$$

Thus, for low-dose tablets, assuming perfect mixing, and a log-normal distribution of particle sizes, the requirement on the coefficient of variation in the distribution of dosage units is essentially a *general requirement on the minimum average number of particles in a dosage unit*. This minimum average number of particles does not depend on label claim.

3.5 Coefficient of skewness and excess for distribution of dose content

In order to assess the validity of the approximation of the distribution of dose content by a normal distribution, the higher moments, i.e. the coefficients of skewness and excess of the distribution of dose content under perfect mixing are of interest.

In (48), the variance, σ_D^2 , in the distribution of dose content was found as

$$\sigma_D^2 = \bar{N}\mu_M^2 \times (C_M^2 + 1).$$

In order to determine the coefficients of skewness and excess consider

$$\kappa_3(D) = \mathbf{E}[(D - \mu_D)^3] = \mathbf{E}_N[\mathbf{E}[(D - \bar{N}\mu_M)^3|N]] \quad (65)$$

$$\begin{aligned} \kappa_4(D) &= \mathbf{E}[(D - \mu_D)^4] - 3\sigma_D^4 \\ &= \mathbf{E}_N[\mathbf{E}[(D - \bar{N}\mu_M)^4|N] - 3\sigma_D^4]. \end{aligned} \quad (66)$$

But

$$\begin{aligned} (D - \bar{N}\mu_M)^3 &= (D - N\mu_M)^3 + 3(D - N\mu_M)^2(N - \bar{N})\mu_M \\ &\quad + 3(D - N\mu_M)(N - \bar{N})^2\mu_M^2 + (N - \bar{N})^3\mu_M^3 \end{aligned} \quad (67)$$

and therefore

$$\begin{aligned}
\mathbf{E}[(D - \bar{N}\mu_M)^3|N] &= \mu_3(D|N) + 3\mu_2(D|N)(N - \bar{N})\mu_M + 0 \\
&\quad + (N - \bar{N})^3\mu_M^3 \\
&= N\mu_3(M) + 3(N - \bar{N})\mu_M N\sigma_M^2 \\
&\quad + (N - \bar{N})^3\mu_M^3 \tag{68}
\end{aligned}$$

$$\mathbf{E}_N[\mathbf{E}[(D - \bar{N}\mu_M)^3|N]] = \bar{N}\mu_3(M) + \mathbf{V}[N]\mu_M\sigma_M^2 + \mu_3(N)\mu_M^3. \tag{69}$$

For the Poisson-distribution $\mathbf{V}[N] = \mu_3(N) = \bar{N}$, and hence

$$\kappa_3(D) = \mathbf{E}_N[\mathbf{E}[(D - \bar{N}\mu_M)^3|N]] = \bar{N}(\mu_3(M) + \mu_M\sigma_M^2 + \mu_M^3). \tag{70}$$

Thus, from (70)

$$\begin{aligned}
\gamma_1(D) &= \frac{\kappa_3(D)}{\sigma_D^3} = \frac{\bar{N}(\mu_3(M) + \mu_M\sigma_M^2 + \mu_M^3)}{\bar{N}^{3/2}\mu_M^3(C_M^2 + 1)^{3/2}} \\
&= \frac{1}{\sqrt{\bar{N}}} \times \frac{\mu_3(M) + \mu_M\sigma_M^2 + \mu_M^3}{\mu_M^3(C_M^2 + 1)^{3/2}}.
\end{aligned}$$

Using (25) we then find

$$\gamma_1(D) = \frac{1}{\sqrt{\bar{N}}} \times \frac{C_M^4 \times (C_M^2 + 3)}{(C_M^2 + 1)^{3/2}} \tag{71}$$

which shows that also in this case, the skewness tends to zero when the number of particles increase.

In order to assess $\mu_4(D)$ consider

$$\begin{aligned}
(D - \bar{N}\mu_M)^4 &= [(D - N\mu_M) + (N - \bar{N})\mu_M]^4 \\
&= (D - N\mu_M)^4 + 4(D - N\mu_M)^3(N - \bar{N})\mu_M \\
&\quad + 6(D - N\mu_M)^2(N - \bar{N})^2\mu_M^2 \\
&\quad + 4(D - N\mu_M)(N - \bar{N})^3\mu_M^3 + (N - \bar{N})^4\mu_M^4
\end{aligned}$$

such that

$$\begin{aligned}
\mathbf{E}[(D - \bar{N}\mu_M)^4]|N &= \mu_4(D|N) + 4\mu_3(D|N)(N - \bar{N})\mu_M \\
&\quad + 6\mu_2(D|N)(N - \bar{N})^2\mu_M^2 + 0 \\
&\quad + \mu_4(N)\mu_M^4 \\
&= N\mu_4(M) + 3N(N - 1)\mu_2(M)^2 + 4N(N - \bar{N})\mu_M\mu_3(M) \\
&\quad + 6N \times \sigma_M^2(N - \bar{N})^2\mu_M^2 + (N - \bar{N})^4\mu_M^4
\end{aligned}$$

and hence

$$\begin{aligned}
\mu_4(D) &= \mathbf{E}_N[\mathbf{E}[(D - \bar{N}\mu_M)^4]|N] \\
&= \bar{N}\mu_4(M) + 3(\mathbf{V}[N] + \bar{N}(\bar{N} - 1))\mu_2(M)^2 + 4\mathbf{V}[N]\mu_M\mu_3(M) \\
&\quad + 6(\mu_3(N) + \bar{N}\mathbf{V}[N])\sigma_M^2\mu_M^2 + \mu_4(N)\mu_M^4.
\end{aligned}$$

For the Poisson distribution we have $\mu_4(N) = 3\bar{N}^2 + \bar{N}$, and therefore

$$\begin{aligned}
\mu_4(D) &= \bar{N}\mu_4(M) + 3\bar{N}^2\mu_2(M)^2 + 4\bar{N}\mu_M\mu_3(M) \\
&\quad + 6\bar{N}(\bar{N} + 1)\sigma_M^2\mu_M^2 + (3\bar{N}^2 + \bar{N})\mu_M^4.
\end{aligned}$$

Using (25) and (26) we obtain

$$\begin{aligned}
\mu_4(D) &= \mu_M^4\bar{N}\left[C_M^4\left((1 + C_M^2)^4 + 2(1 + C_M^2)^3 + 3(1 + C_M^2)^2 - 3\right)\right. \\
&\quad \left.+ 3\bar{N}C_M^4 + 4C_M^4(C_M^2 + 3)\right. \\
&\quad \left.+ 6(\bar{N} + 1)C_M^2 + (3\bar{N} + 1)\right] \\
&= 3\mu_M^4\bar{N}^2(1 + C_M^2)^2 \\
&\quad + \mu_M^4\bar{N}\left[C_M^4\left((1 + C_M^2)^4 + 2(1 + C_M^2)^3 + 3(1 + C_M^2)^2 - 3\right)\right. \\
&\quad \left.+ 4C_M^4(C_M^2 + 3) + 6C_M^2 + 1\right]
\end{aligned}$$

such that we obtain

$$\gamma_2(D) = \frac{1}{\bar{N}} \times \frac{C_M^4\left((1 + C_M^2)^4 + 2(1 + C_M^2)^3 + 3(1 + C_M^2)^2 - 3\right) + 4C_M^4(C_M^2 + 3) + 6C_M^2 + 1}{(C_M^2 + 1)^2} \quad (72)$$

which shows that the excess of the distribution of the doses tends to zero when the number of particles increase.

Thus, the effect of the deviation from the normal distribution of particle sizes becomes less pronounced the larger the average number of particles.

Invoking (45) to express the average number of particles in terms of the label claim, (71) may be expressed as

$$\gamma_1(D) = \sqrt{\frac{\mu_M}{LC}} \times \frac{C_M^4 \times (C_M^2 + 3)}{(C_M^2 + 1)^{3/2}} \quad (73)$$

and similarly, (72) yields

$$\gamma_2(D) = \frac{\mu_M}{LC} \times \frac{C_M^4 \left((1 + C_M^2)^4 + 2(1 + C_M^2)^3 + 3(1 + C_M^2)^2 - 3 \right) + 4C_M^4(C_M^2 + 3) + 6C_M^2 + 1}{(C_M^2 + 1)^2} \quad (74)$$

4 Distribution of dose content under non-random mixing

We still consider the model (33) for the dose, D .

However, instead of the random dispersion of dose particles over the blend modelled by a distribution of the number of particles in the blend as a homogeneous Poisson point process in the three-dimensional space, we shall consider a model for a nonhomogeneous blend.

The homogeneous Poisson point process model is naturally extended to a Poisson point process where the intensity is varying over the blend. One such extension is the *doubly stochastic Poisson Process*, or *Cox Process*, see e.g. Stoyan et al [65]. Such a process can be thought of as arising from a two-step random mechanism. The first step generates a random intensity measure over the blend, and the second step generates a Poisson process corresponding to that intensity measure.

Let $\Lambda(\cdot)$ denote the random intensity measure defined over the blend, and let

$$\Lambda_v = \int_{u \in v} \Lambda(u) du$$

denote the random volume-intensity corresponding to a volume v . Analogously let $\lambda(\cdot)$ denote a realization of this measure with

$$\lambda_v = \int_v \lambda(u) du$$

denoting the realized volume-intensity in a volume v .

When a sample of a fixed volume, v , is selected from the blend, the sampling process may similarly be thought of as a two-step random mechanism. First, a random intensity Λ_v corresponding to that volume is selected from a distribution of volume-intensities, and secondly a number of particles is generated as a Poisson distributed random variable with mean λ_v corresponding to that intensity.

In order to obtain a mathematically tractable form of the resulting distributions, it will be assumed that the doubly stochastic Poisson process is such that the distribution of Λ_v under repeated sampling from the blend may be described by a gamma distribution with mean

$$\mathbf{E}[\Lambda_v] = v \times \nu \quad (75)$$

and with variance

$$\mathbf{V}[\Lambda_v] = \mathbf{E}[\Lambda_v] \times \gamma_v^2 = v \times \nu \times \gamma_v^2 \quad (76)$$

where ν denotes the overall average number of particles in the blend per unit volume ($[\mu\text{m}^3]$) and where the measure, γ_v , of the clustering tendency is related to the coefficient of variation in the distribution of λ_v ,

$$C_\Lambda = \frac{\sqrt{\mathbf{V}[\Lambda_v]}}{\mathbf{E}[\Lambda_v]} = \frac{\gamma_v}{\sqrt{v \times \nu}}. \quad (77)$$

The suffix v on γ_v^2 serves to emphasize that the variance in the distribution of Λ_v may depend on the sample amount v . Moreover, γ_v^2 depends on the second

order (spatial correlation) properties of the random intensity measure $\Lambda(\cdot)$: one extreme being complete independence between intensities in infinitesimal neighbouring volumes in which case γ_v^2 does not depend on the sample amount v , i.e. $\gamma_v^2 = c$. Under a perfect random mixing, i.e. a homogeneous Poisson process with (constant) intensity ν one has $\gamma_v^2 = 0$.

It will moreover be assumed that the spatial correlation corresponding to the intensity measure $\Lambda(\cdot)$ fades away so rapidly that the volume-intensities, Λ_v , corresponding to independently sampled volumes, v , may be approximated by independent samples from the above distribution.

Formally, it is assumed that the conditional distribution of the number, N , of particles in a random volume, v , given the volume intensity $\Lambda_v = \lambda_v$ corresponding to that volume may be described by a Poisson distribution with mean λ_v , and that the volume-intensities corresponding to the volumes sampled may be described by independent identically gamma-distributed variables Λ_v with mean (75) and variance (76).

Under this assumption it may be shown (see e.g. Johnson and Kotz [66]) that the *marginal distribution* of the number of particles in a sample of volume v is a negative binomial distribution with mean and variance

$$\bar{N} = \mathbf{E}[N] = v \times \nu \quad (78)$$

$$\mathbf{V}[N] = \mathbf{E}_\Lambda[\mathbf{V}[N|\Lambda_v]] + \mathbf{V}_\Lambda[\mathbf{E}[N|\Lambda_v]] \quad (79)$$

$$= \mathbf{E}[\Lambda_v] + \mathbf{V}[\Lambda_v] = v\nu(1 + \gamma_v^2) \quad (80)$$

$$= \mathbf{E}[N] \times (1 + \gamma_v^2) \quad (81)$$

i.e. γ_v^2 expresses the *overdispersion* of counts as compared with the Poisson distribution. For $\gamma_v = 0$ the negative binomial distribution of particle counts degenerates to a Poisson distribution with mean $\bar{N} = v\nu$.

The *intracluster correlation* in volumes of size v is

$$\rho_v = \frac{\gamma_v^2}{1 + \gamma_v^2} = \frac{1}{1 + 1/\gamma_v^2}. \quad (82)$$

Assuming further that particle sizes vary independent of the volume-intensity, i.e. particles in clusters are selected independently from the distribution of

particle sizes, then the moments in the conditional distribution of doses given the number of particles in the tablet are given by (34) to (37).

It then follows from (46) and (47), that the (marginal) mean and variance in the distribution of doses are

$$\mu_D = \mathbf{E}[D] = \bar{N}\mu_M = v \times \nu \times \mu_M \quad (83)$$

$$\sigma_D^2 = \mathbf{V}[D] = \bar{N}\sigma_M^2 + \mu_M^2\mathbf{V}[N] \quad (84)$$

$$\begin{aligned} &= \bar{N} \times (\sigma_M^2 + \mu_M^2(1 + \gamma_v^2)) \\ &= \bar{N}\mu_M^2 \times (C_M^2 + 1 + \gamma_v^2) \end{aligned} \quad (85)$$

and

$$C_D = \frac{1}{\sqrt{\bar{N}}} \sqrt{1 + C_M^2 + \gamma_v^2}. \quad (86)$$

Under the homogeneity (Poisson) assumption we have $\mathbf{V}[N] = \mathbf{E}[N] = \bar{N}$, and in that case the expression for the variance simplifies to (48).

In the general case the coefficient of variation in the distribution of doses is

$$C_D = \frac{\sigma_D}{\mu_D} = \frac{1}{\sqrt{\bar{N}}} \times \sqrt{1 + C_M^2 + \gamma_v^2} \quad (87)$$

(with (49) as a special case for $\gamma_v = 0$).

Comparing (87) and (49) for a given number distribution of particle masses, it is seen that the overdispersion of the distribution of the number of particles under imperfect mixing implies an overdispersion of doses as compared to the distribution of doses under perfect mixing (Poisson distribution of number of particles).

Expressed in terms of label claim (LC) one finds using (45)

$$\sigma_D^2 = \mathbf{V}[D] = \text{LC} \times \mu_M (C_M^2 + 1 + \gamma_v^2) \quad (88)$$

$$C_D = \sqrt{\frac{\mu_M (C_M^2 + 1 + \gamma_v^2)}{\text{LC}}}. \quad (89)$$

Consequently, when relative doses, $d = D/\text{LC}$ are considered, the variance, σ_d^2 of the distribution of relative doses is

$$\sigma_d^2 = \text{V}[D/\text{LC}] = C_D^2 = \frac{\mu_M}{\text{LC}} (C_M^2 + 1 + \gamma_v^2) \quad (90)$$

with the coefficient of variation, C_d , in the distribution of relative doses given by (89) as $C_d = C_D$.

4.1 Spherical particles

For spherical particles, μ_M and C_M are given by (21) and (22) and one obtains

$$\sigma_D^2 = \bar{N} \mu_M^2 \times (C_M^2 + 1 + \gamma_v^2) = \bar{N} \mu_M^2 \times \left((1 + C_R^2)^9 + \gamma_v^2 \right) \quad (91)$$

$$= \bar{N} \times \left[(4\pi/3) \rho \mu_R^3 \times (1 + C_R^2)^3 \right]^2 \times \left((1 + C_R^2)^9 + \gamma_v^2 \right). \quad (92)$$

The coefficient of variation in the distribution of doses is

$$C_D = \frac{\sigma_D}{\mu_D} = \frac{1}{\sqrt{\bar{N}}} \times \sqrt{(1 + C_R^2)^9 + \gamma_v^2} \quad (93)$$

(with (58) as a special case for $\gamma = 0$).

Comparing (93), (58) and (40) for a given number distribution of particle radii, it is seen that the overdispersion of the distribution of the number of particles (as compared to the Poisson distribution) under imperfect mixing implies an overdispersion of doses as compared to the distribution of doses under perfect mixing.

Expressed in terms of label claim (LC) one finds using (45)

$$\sigma_D^2 = \text{V}[D] = \frac{4\pi \text{LC} \times \rho}{3} \times \mu_R^3 (1 + C_R^2)^3 \left((1 + C_R^2)^9 + \gamma_v^2 \right) \quad (94)$$

$$C_D = \sqrt{\frac{4\pi \rho \mu_R^3 (1 + C_R^2)^3}{3 \text{LC}}} \times \left((1 + C_R^2)^9 + \gamma_v^2 \right) \quad (95)$$

which shows that under the assumption of perfect mixing and a given distribution of particle sizes does the absolute standard deviation increase proportional

to the square root of the magnitude of the dose (LC), and hence the coefficient of variation decreases proportional to the square root of the dose. The larger the dose, the smaller a coefficient of variation in the distribution of doses.

Consequently, when relative doses, $d = D/LC$ are considered, the variance, σ_d^2 of the distribution of relative doses is

$$\sigma_d^2 = V[D/LC] = \frac{4\pi \times \rho}{3LC} \times \mu_R^3 (1 + C_R^2)^3 ((1 + C_R^2)^9 + \gamma_v^2) \quad (96)$$

with the coefficient of variation, C_d , in the distribution of relative doses given by (95) as $C_d = C_D$.

4.2 Minimum number of particles necessary to secure a specified dose coefficient of variation

In analogy with Section 3.4 consider a requirement $C_D \leq C_{max}$. Under the model for nonrandom mixing, the resulting requirement on the average number of particles is found from (93)

$$\bar{N} \geq \frac{(1 + C_R^2)^9 + \gamma_v^2}{C_{max}^2}. \quad (97)$$

In analogy with (63) the requirement $C_D \leq C_{max}$ may be expressed as a requirement to the mean particle radius (for given LC, C_R and γ_v^2) as

$$\mu_R \leq \frac{[C_{max}^2 \times 3LC/(4\pi\rho)]^{1/3}}{(1 + C_R^2) \times [(1 + C_R^2)^9 + \gamma_v^2]^{1/3}}. \quad (98)$$

Thus, also in this case the limit for acceptable particle radii increases proportional to the third root of LC, and decreases with increasing coefficient of variation, C_R , for the distribution of particle radii, although in a less simple manner as in (63).

5 Discussion

In production of tablets, variation is introduced at various stages in the process:

sieving/milling process: the purpose of the sieving/milling of the drug particles is to reduce the particles to an adequate size. Natural variation in process output is described by a unimodal distribution, often a normal, or a log normal distribution.

blending/mixing process: a purpose of the blending/mixing is to obtain a uniform mixture of drug particles with other components. Natural variation in a satisfactory blend might be modelled by a random spatial distribution of drug particles.

transfer to tablet press and tableting process: ideally, the transfer aims at transferring the spatial distribution in the blend at tablet size level.

In the paper we have investigated the transfer of variation resulting from the sieving/milling process through the mixing process to samples from the blend, or to tablets in ideal situations where only natural variation is present, i.e. no sampling bias, no deblending etc., (only some clustering tendencies in the blend have been considered).

In Section 2 properties derived from a lognormal distribution of particle radii were discussed with the aim of assessing the departure from normality of the resulting distribution of particle mass. It was found that the relative variation (the coefficient of variation) in the distribution of radii is amplified when considering the distribution of particle mass. As the coefficient of variation in a lognormal distribution is determinant for the departure from normality (skewness and excess), the departure from normality of the distribution of particle mass is heavily dependent on the magnitude of the coefficient of variation in the distribution of radii.

The variation in dose content of tablets is caused by variation in the number of dose particles in the tablets as well as variation in the mass of individual particles. In Section 3.2 the contribution from the variation in particle mass has been investigated keeping the number of particles fixed. Expressions for the moments in the distribution of dose content have been derived and expressed in terms of label claim, mean particle size and coefficient of variation in the distribution of particle radii. Alternative expressions in terms of number of particles and coefficient of variation in the distribution of particle radii have

been provided. It was found that the effect of skewness and excess in the distribution of particle masses is diminished when particle masses are added to form tablets. The larger the number of particles, the smaller the skewness and excess in the distribution of doses, and the better will the distribution of doses be approximated by a normal distribution.

Moreover, it was found that the coefficient of variation in the distribution of dose content (and of relative content) is inversely proportional to the square root of the number of particles. Thus, the smaller the number of particles, the larger a variance in the distribution of relative doses. Therefore, in order for low dose tablets to comply with usual dosage uniformity test, it is imperative that particle diameter is small, and it is furthermore advantageous that the coefficient of variation in the distribution of particle sizes is suitably small.

In Section 3.3 the further variation introduced by the variation of the number of particles in the tablets has been investigated under the assumption of a homogeneous blend, i.e. a random spatial distribution of dose particles in the blend. Expressions for the moments in the distribution of dose content have been provided. As could be expected, the variation of the number of particles amplifies the variation in dose due to varying particle sizes. However, unless the distribution of particle masses is very narrow, the major contribution to the variation in dose content originates from the variation in particle masses, and hence, also in this case the coefficient of variation in the distribution of dose content is inversely proportional to the square root of the number of particles.

Thus, assuming perfect mixing, and a log-normal distribution of particle sizes, the requirement on the coefficient of variation in the distribution of dosage units is essentially a *general requirement on the minimum average number of particles in a dosage unit*. This minimum average number of particles does not depend on label claim. However, as the average number of particles in tablets depend on label claim (for a given particle size distribution), a blend that might produce a satisfactory distribution of doses (in terms of the coefficient of variation in the distribution of relative doses) for large dose tablets need not be satisfactory for smaller dose tablets.

In Section 3.4, explicit expressions for the minimum average number of particles per tablets necessary to secure a specified coefficient of variation in the distribution of doses have been provided under the assumption of a homogeneous blend, and in Section 3.5 expressions for coefficients of skewness and excess for the distribution have been provided.

Finally, in Section 4, the results are extended to cover also such nonrandom mixtures that may be modelled by a hierarchical model representing a spatial variation of the intensity of particles.

Results have been derived under the ideal assumption of spherical particles. Therefore, for practical use, a “nonsphericity factor” might be introduced in the transformation from distribution of particle radii to particle mass and vice versa. In the paper we have allowed for this possibility by providing separate expressions for moments in the distribution of doses in terms of properties of the distribution of particle masses, valid regardless of shapes, supplemented by expressions in terms of properties of the distribution of particle radii, valid only under the further assumption of spherical particles.

Interpreting the results in terms of blend samples rather than samples of tablets from a batch, it is of interest to note that the practical necessity of using blend samples that are larger than the dosage units imply that the coefficient of variation in such blend samples is smaller than the coefficient of variation in the resulting dosage units. For blend samples that are four times the size of the final dosage units, the coefficient of variation in the blend samples is only half the size of the coefficient of variation in the final dosage units. Moreover, a larger blend sample might mask departure from normality in the distribution of dose content in low-dose tablets.

6 List of symbols

Symbol	Units	
$E[X]$		Mean of the distribution of X
$V[X]$		Variance of the distribution of X
$E_N[X]$		Mean (under variation of N) of the distribution of X
$V_N[X]$		Variance (under variation of N) of the distribution of X
$\mu_3(X)$		Third moment about the mean in the distribution of X $\mu_3(X) = E[(X - E[X])^3]$
$\mu_4(X)$		Fourth moment about the mean in the distribution of X $\mu_4(X) = E[(X - E[X])^4]$
$\gamma_1(X)$	-	Coefficient of skewness of the distribution of X (fraction) - $\gamma_1(X) = \mu_3(X)/(V[X])^{3/2}$
$\gamma_2(X)$	-	Coefficient of excess of the distribution of X (fraction) - $\gamma_2(X) = \mu_4(X)/(V[X])^2 - 3$
R	μm	Radius of a particle
μ_R	μm	Mean particle radius
σ_R	μm	Standard deviation in particle radius distribution
C_R	-	Coefficient of variation in particle radius distribution (fraction) $C_R = \sigma_R/\mu_R$
β_R		Standard deviation in distribution of log particle radius
α_R		log(Median) in particle radius distribution
V	μm^3	Particle volume
ρ	$\mu\text{g}/\mu\text{m}^3$	Mass density of particles
M	μg	Particle mass
μ_M	μg	Mean particle mass
σ_M	μg	Standard deviation in particle mass distribution
C_M	-	Coefficient of variation in particle mass distribution (fraction) $C_M = \sigma_M/\mu_M$
N		Number of particles in tablet
\bar{N}		Average number of particles in tablet
D	μg	Dose (Total mass of drug in tablet) $D = \sum_{i=1}^N M_i$
μ_D	μg	Mean dose
σ_D	μg	Standard deviation in dose distribution
C_D	-	Coefficient of variation in dose distribution (fraction) $C_D = \sigma_D/\mu_D$
LC	μg	Required dose, Label Claim
d	-	Relative dose (fraction) $d = D/\text{LC}$
μ_d	-	Mean relative dose
σ_d	-	Standard deviation in distribution of relative doses

Paper F

Case: Analysis of homogeneity in production scale batches

1 Purpose

This case relates to the sampling from a low dose ($25\mu g$) product. An experimental design was set up with the purpose to investigate:

1. whether the type of thief shown to the left in Figure 2.3 in Chapter 2 is qualified as sampling device for the product. Tablet samples from the same batch are used to evaluate this problem.
2. bias (difference in mean content) and repeatability of samples sampled by respectively different thieves, persons and thief tips.
 - (a) whether samples taken with respectively a long and a short thief lead to the same result (it might be more difficult to handle a long thief, or the samples taken with the long thief may be biased because they are primarily sampled from the bottom of the batch).
 - (b) whether the size (1X or 3X unit dose, 80 mg) influences the result.
 - (c) whether the sampling procedure is sensitive of the person collecting the sample.
3. the homogeneity of each batch.
 - (a) whether the three layers in a batch lead to equally representative samples.

2 Experimental

The original experimental design included twenty-six blend samples, fifty tablet samples and a larger blend sample for sieving analysis from each of three batches of the product.

However, blend samples were taken from the first two batches, but the third batch was transferred to the tablet press before the blend samples were taken.

From each of the first two batches ten tablet samples were collected after one hour of tableting.

In total 26 samples were collected from each of two batches as shown in Table 1.

Nr.	Batch	Layer	Area	Theif	Pers.	Nr.	Batch	Layer	Area	Theif	Pers.
1	1	Top	C	1	X	1	2	Top	A	1	Y
2	1	Top	C	1	X	2	2	Top	A	1	Y
3	1	Top	B	1	X	3	2	Top	C	1	Y
4	1	Top	B	1	Y	4	2	Top	B	1	X
5	1	Top	A	1	X	5	2	Top	D	1	X
6	1	Top	D	1	Y	6	2	Top	D	1	Y
7	1	Middle	B	2	X	7	2	Middle	D	3	Y
8	1	Middle	B	2	X	8	2	Middle	D	2	Y
9	1	Middle	B	3	X	9	2	Middle	D	2	Y
10	1	Middle	C	3	Y	10	2	Middle	A	3	X
11	1	Middle	C	2	Y	11	2	Middle	A	1	X
12	1	Middle	C	1	Y	12	2	Middle	A	3	X
13	1	Middle	D	3	Y	13	2	Middle	B	2	Y
14	1	Middle	D	1	Y	14	2	Middle	B	1	Y
15	1	Middle	D	3	Y	15	2	Middle	B	1	Y
16	1	Middle	A	1	X	16	2	Middle	C	1	X
17	1	Middle	A	1	X	17	2	Middle	C	3	X
18	1	Middle	A	2	X	18	2	Middle	C	2	X
19	1	Bottom	D	3	Y	19	2	Bottom	B	2	Y
20	1	Bottom	C	3	Y	20	2	Bottom	B	2	Y
21	1	Bottom	C	2	Y	21	2	Bottom	B	3	Y
22	1	Bottom	A	2	X	22	2	Bottom	B	3	Y
23	1	Bottom	A	2	X	23	2	Bottom	A	3	X
24	1	Bottom	A	3	X	24	2	Bottom	D	2	Y
25	1	Bottom	A	3	X	25	2	Bottom	C	2	X
26	1	Bottom	B	2	X	26	2	Bottom	C	3	X

Table 1: The table shows the samples collected from two blend batches.

Samples were collected from three layers, and four areas within each layer by two persons using three different sampling thieves.

Regarding areas, four areas are identified in each layer as shown in Figure 1. In the middle layer the orientation of the areas are rotated 60 degrees compared to the top and the bottom layer. If more than one sample is taken in the same area

and layer, the samples are sampled as close as possible - but not from exactly the same position. Estimates of the small scale variation in the blend are based on these samples.

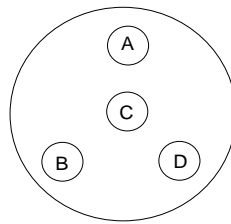


Fig. 1: A sectional plane of a layer in the blend. Each layer contains four areas: A, B, C and D.

The three different thieves used for sampling are:

- Thief 1 is a short thief with a 1X dose tip.
- Thief 2 is a long thief with a 1X dose tip.
- Thief 3 is a long thief with a 3X dose tip.

The same two persons are sampling in the two batches. Person X is a person with above average skills in sampling techniques.

Because of several physical restrictions the blend samples are not collected in a completely randomized order.

3 Statistical Analysis

The analyses are performed in SAS using procedures for Generalized Linear Models (GENMOD) and General Linear Mixed Models (MIXED).

The actual collected blend samples correspond to a design with confounding between factors. Further only tablets corresponding to the bottom of the blend have been collected. The results of the statistical analysis is therefore only a guide of the conditions in these two batches.

The statistical analysis is performed on the relative dose, i.e. content of the active component divided by label claim.

The mean relative content is 0.912 in blend samples from batch 1 and 1.016 in blend samples from batch 2. The relative dose in the individual samples are shown in Figure 2.

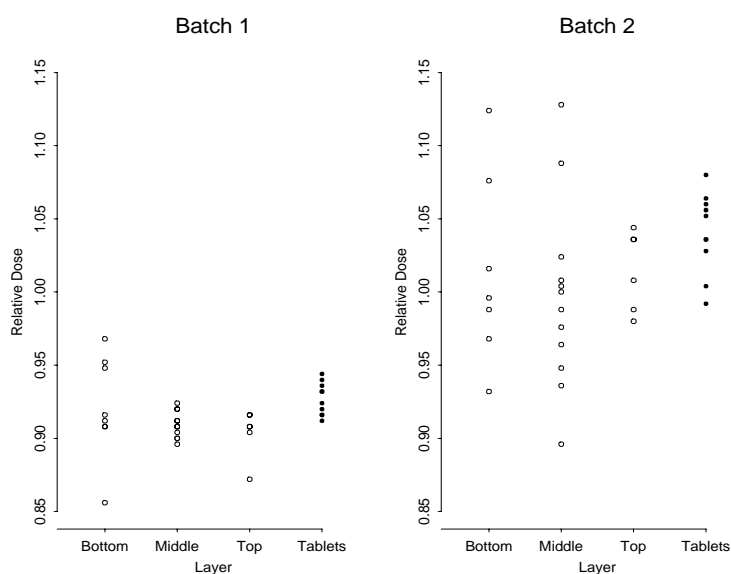


Fig. 2: Results from all blend- and tablet samples.

3.1 Assessment of tablet samples

Summary results for the tablets are given in Table 2.

Figure 2 indicates that the relative content in samples from batch 1 is less than the relative content in samples from batch 2. A Welch modified two-

	Dose (μg)		Relative Dose		
	\bar{X}	S^2	\bar{X}	S^2	Conf. Int. S^2
Batch 1	23.18	0.0773	0.9272	0.0001	[0.0001;0.0004]
Batch 2	26.02	0.4684	1.0408	0.0007	[0.0004;0.0025]

Table 2: Mean, variance and 95% confidence interval for the dose and the relative dose for tablet samples.

sample t-test on the tablet samples shows a significant difference between the two samples. The plot also reveals that in general the variation in batch 2 is larger than the variation in batch 1.

Figure 3 shows QQ-plot for the tablets from each batch. The content of active component in the tablets may be considered to be normally distributed, with mean and variance depending on the batch from which the tablets are sampled.

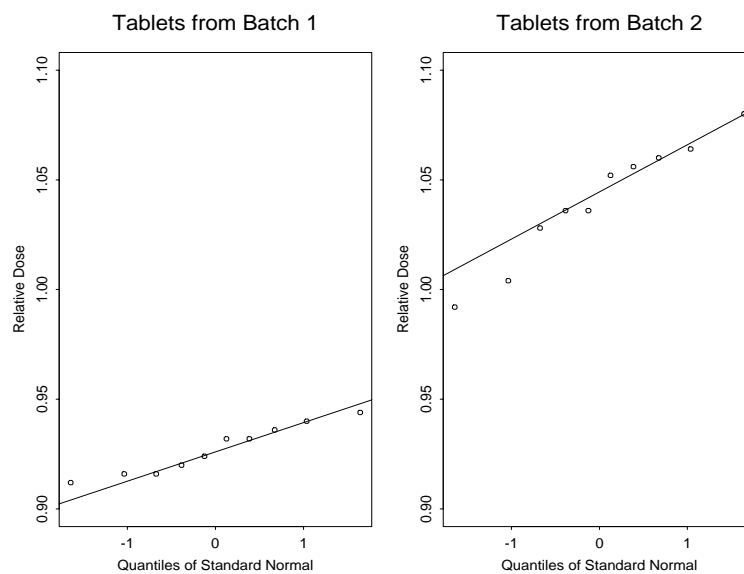


Fig. 3: QQ-plot for ten tablets sampled from each of the two batches. The tablets were sampled when the tablet press had been running for a period of 1 hour. The Normal distribution gives a satisfactory fit of the data.

An F-test shows that the variation between tablets from batch 2 is significantly larger than the variation between the tablets from batch 1. This indicates that

batch 1 is more homogeneous i.e. that the small scale variation in batch 1 is less than the small scale variation in batch 2.

3.2 Repeatability

The repeatability variance is the variance between replicate samples from the same batch, layer and area by the same person and sampling thief. The repeatability variance incorporates contributions from small scale variation in the blend, variation due to sampling and due to the chemical analysis. Further, as blend samples in contrast to tablet samples are adjusted for the weight of the sample, uncertainty on the weight measurement also has an influence on the repeatability variance between blend samples.

With 6 pairs of replicates in each batch it has been possible to assess the effect of different sampling thieves and layers on the repeatability variance. The number of actual collected samples is too small to include the effect of persons in this statistical analysis. Figure 4 indicates that the overall variation between samples collected by each person do not differ, indicating that the persons are equally consistent in the way they take a sample.

The statistical model for the repeatability variance is a Generalized Linear Model:

$$\ln \sigma_{layer,thief}^2 = \mu^* + \alpha_{layer}^* + \beta_{thief}^*,$$

$$\text{where } S_{layer,thief}^2 \in \sigma_{layer,thief}^2 \chi^2(1).$$

The results are discussed below.

Thieves

The statistical analysis shows that use of different sampling thieves have no significant influence on the repeatability variance. This means that the repeatability variance introduced by the three thieves do not differ significantly. How-

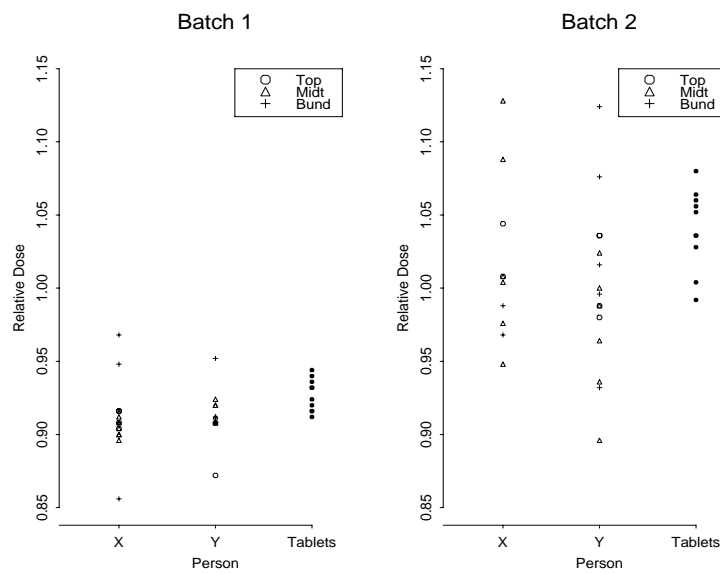


Fig. 4: Variation between replicate samples collected by each of the two persons. This variation includes variation due to variations in the blend, variation among sampling thieves, analytical error and variation from adjusting the blend samples for the weight. Results from the tablet samples are included as reference in the plot.

ever, the tendency for both batch 1 and batch 2 is that using sampling thief 1 introduces less sampling error than using thief 3. As seen in Table 3 the estimated standard deviation between samples collected by thief 3 is approximate four to five times as big as the estimated standard deviation between samples collected by thief 1. It is conceivable that an analysis of the results from an experimental design with more experiments will find the tendency significant.

Thief	Batch 1 Middle layer	Batch 2 Middle layer
1	$(0.0028)^2$	$(0.0170)^2$
2	$(0.0064)^2$	$(0.0495)^2$
3	$(0.0103)^2$	$(0.0831)^2$

Table 3: Variation between samples collected with various thieves in the middle layer in batch 1 and batch 2.

Layers

The repeatability variance is not found to differ significantly in the three layers in a batch. However for both batches the tendency is that the variation between replicates is less in the top, than in the bottom layer. This can be interpreted as the batches not being homogeneous on the small scale level. It is conceivable that the tendency will be significant in an experimental design with more observations.

Comparing variation in the blend with variation between tablets

Ten tablets from each batch were sampled as closely as possible in time after one hour of tableting, i.e. at the beginning of the tableting process. The variation between the tablets from a batch includes the variation in the blend, the variation introduced by the tablet press (deblending and weight variation) and the chemical analysis. However, it is assumed, that the weight variation and the variation due to deblending introduced at the tablet press are negligible and therefore the variation between tablets can be regarded as an estimate of

the small scale variation in the blend, i.e. a measure of the minimum obtainable variation between samples from the blend. It is therefore relevant to compare the variation among tablets to variations in the blends.

The problem is to decide which variations in the blends should be compared to the variation between tablets.

Even though not statistically significant the tendency is that the variation between replicates in the top layer is smaller than the variation between replicates in the bottom layer. Therefore, comparing variation between tablets to variation between replicates in respectively the top and the bottom layer may not lead to the same result regarding an estimate of the small scale variation in the blend.

To assess whether an estimate of the small scale variation in the blend should be based on the top or the bottom layer, assumptions have to be made about a physically explanation for the tendency for the variation between replicates to be smaller in the top layer than in the bottom.

One explanation is that the real small scale variation in the blend differs among layers. Another explanation is that it is more difficult to hit the same spot in a lower layer of a blend when sampling replicates. It is not possible due to the experimental design to decide which explanation is correct.

The two physically explanations lead to the following estimates of the small scale variation in the blend:

1. **Difficulties with hitting the same spot is the main reason for the tendency of the variation between replicates being smaller in the top than in the bottom of a batch.** The best estimate of the small scale variation in the blend is an estimate based on replicates sampled with the thief that introduces the smallest sampling variance, i.e thief 1. The estimate is based on the top layer because it is easier to hit the same spot in this layer than in the middle and the bottom layer. A 95% confidence interval for the small scale variation, $\sigma_{small\ scale}^2$, in batch 1 is [0.0000;0.0001] and [0.0000;0.0005] for batch 2.

2. **The real small scale variation is not constant among the three layers in the batch.** As the tablets are sampled at the beginning of the tableting process, they correspond to the bottom layer of the blend. Therefore the best estimate of the small scale variation in the blend corresponding to the variation among the collected tablets is based on the replicates sampled with thief 2 in the bottom of the blend. Thief 2 is chosen because it is the thief used in the bottom layer that introduces the smallest sampling variance. A 95% confidence interval for the small scale variation in the bottom layer is [0.0001;0.0063] for batch 1 and [0.0004;0.0272] for batch 2.

95% confidence interval for the small scale variation based on the tablets are [0.0001;0.0004] for batch 1 and [0.0004;0.0025] for batch 2. Confidence intervals for the small scale variation based on the tablets are more overlapping to the confidence interval corresponding to situation 2 than the confidence interval corresponding to situation 1.

If the small scale variation estimated from the tablets represents the true small scale variation in the blend, the small scale variation estimated from blend samples must be either of the same size or (when variation due to sampling is not negligible) larger. However, in situation 1 the estimate based on blend samples is smaller than the estimate based on tablet samples, thus implying that the true small scale variation in the bottom layer actually is larger than in the top layer.

3.3 Mean content of the active component

The statistical model for the relative content in a sample from a given batch is

$$y_{layer,area,person,thief} = \mu + \alpha_{layer} + \beta_{thief} + \gamma_{person} + D_{area} + E_{layer,area,person,thief},$$

where $D_{area} \in N(0, \sigma_{area}^2)$ and $E_{layer,area,person,thief} \in N(0, \sigma_E^2)$.

The model is a General Linear Mixed Model. Under this model it has been tested whether the mean content of the active component differs in the three

layers in a blend and whether the mean content in a sample depends on the thief and the person collecting the sample.

The residual variance from this analysis is an estimate of the variance between replicates in the batch. The estimates are 0.000227 for batch 1 and 0.005141 for batch 2. These estimates are average of the variance between replicates regardless of the actual person, thief, layer and area from which the sample is taken. Estimates of the small scale variation (variation between replicates) from the analysis of repeatability (Section 3.2) take into account the the thief used to collect the sample and the layer from which the sample is collected.

The results are discussed in more detail in the following.

Layers

For both batches the tendency is that the mean content in the top layer is less than the mean content in the bottom layer.

The analysis also gives an estimate for the variance between the mean content of areas within a layer. For batch 1 the estimate is $0.00036 = (0.01897)^2$. For batch 2 it has not been possible to distinguish the variation between areas from the variation between replicates within an area.

Thieves

Thief 1 is a short thief used in the top and the middle layer, whereas both thief 2 and 3 are long thieves used in the middle and the bottom layer. Thief 1 and thief 2 collect samples of the size 1X dose and thief 3 collect samples of the size 3X dose. Thus, the tip used to thief 1 and thief 2 are similar but not identical. Because both thief 2 and 3 are long thieves they are interchangeable and for future sampling it is relevant to know which thief to choose.

In both batches the tendency is that the mean content of the active component is less in samples collected with thief 2 than in samples collected with thief 3.

The difference is significant on the 10% level in batch 1 and 16% level in batch 2.

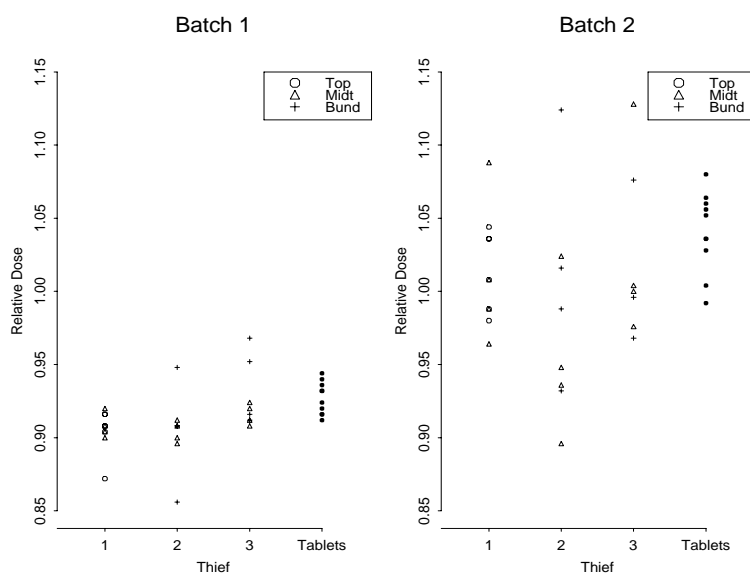


Fig. 5: In general the relative content in samples collected with thief 2 is smaller than the relative dose in the tablets.

Regarding content the tablet samples are in general more similar to samples collected by thief 3 than to samples collected by thief 2. This indicates that thief 2 introduces bias. The fact that thieves may introduce bias causing the content in blend samples to differ from the content in tablet samples is in agreement with the literature [5] as well as with Figure 2.4 on page 12. The literature (e.g. [21]) also gives several examples that the risk of bias is larger when the blend sample is small, i.e. close to unit dose. This is in agreement with thief 3 taking samples of 3X dose, whereas thief 2, which is under suspicion of sampling bias, takes samples of 1X dose.

In batch 1 the estimated bias on the mean content in samples collected with thief 2 is -0.015 ± 0.0184 (95% confidence interval) relatively to the mean content in samples collected with thief 3. That is, if the relative content in a sample collected with thief 3 is 0.98, the relative content would have been 0.965 ± 0.0184 had thief 2 been used to collect the sample.

In batch 2 the estimated bias is -0.057 ± 0.0814 (95% confidence interval), i.e. a relative content of 0.98 in a sample collected with thief 3 corresponds to a relative content of 0.923 ± 0.0814 had thief 2 been used.

Conclusion

Homogeneity of the batches

From blend- and tablet samples it appears that the mean content in batch 1 is less than the mean content in batch 2, and that in general the variation in batch 2 is larger than the variation in batch 1.

The mean content in the three layers in a batch does not differ significantly. However, there is a tendency that the small scale variation in a top layer is smaller than the small scale variation in the bottom.

In batch 1 the variation between replicates (small scale variation) are smaller than the variation between areas within a layer. In batch 2 the variation between replicates is so large that it has not been possible to separate the variation between replicates from the variation between areas within a layer.

It is concluded, that batch 1 is more well mixed than batch 2.

Persons

The mean content in samples collected by person 1 does not differ significantly from the mean content in samples collected by person 2. It has not been possible to test whether sampling error (variation) introduced by the two persons is of the same order of magnitude, but a plot indicates that neither person introduces significantly larger sampling error than the other.

It can not be rejected that the persons have similar skills in sampling.

Thieves

The tendency is that thief 3 introduces the largest sampling error (variation) and that thief 1 introduces the smallest sampling error. A possible explanation is that it is easier to handle a short thief than a long thief.

Even though thief 2 generally introduces less sampling error than thief 3, thief 3 is recommended for sampling in the bottom layer, as thief 2 is under suspicion of sampling bias.

It is noteworthy that the long thief with the small chamber in the tip (thief 2, 1X dose) is under suspicion of sampling bias, whereas there is no indication of sampling bias for the long thief with the large chamber in the tip (thief 3, 3X dose). This conclusion is based on a comparison of variation between replicates from the blends with variation between tablets.

Bibliography

- [1] Muzzio, F. J., Roddy, M., Brone, D., Alexander, A. W., and Sudah, O. An improved Powder-Sampling Tool. *Pharmaceutical Technology*, pages 92–110, April 1999.
- [2] Berman, J. The Compliance and Science of Blend Uniformity Analysis. *PDA Journal of Pharmaceutical Science & technology*, 55(4), 2001.
- [3] United States of America vs. Barr Laboratories Inc. Civil action for the District of New Jersey, February 1993.
- [4] Center for Drug Evaluation FDA and Research. 21 Code of Federal Regulations Parts 210 and 211.
- [5] Berman, J. and Planchard, J. A. Blend Uniformity and Unit Dose Sampling. *Drug Development and Industrial Pharmacy*, 21(11), 1995.
- [6] FDA. The Food and Drug Administration: An Overview. Publication No. BG99-2, www.fda.gov, January 1999.
- [7] USP. About USP. www.usp.org, 2000.
- [8] PDA. About PDA. www.pda.org, 2000.
- [9] PQRI. About PQRI. www.pqri.org, 2000.
- [10] Department of Health and FDA Human Services. 21 CFR Parts 210 and 211 Current Good Manufacturing Practice: Amendment of Certain Requirements for Finished Pharmaceuticals; Proposed Rule. In *Federal Register, Proposed Rules*, volume 61, pages 20103 – 20115. FDA.

- [11] Draft Guidance for Industry ANDAs: Blend Uniformity analysis, August 1999.
- [12] Schilling, E. G. *Acceptance Sampling in Quality Control*. Marcel Dekker, Inc., 1982.
- [13] Kristensen, H. G. The Characterization of Non-Random Mixtures. A survey. *Powder Technology*, 13:103–113, 1976.
- [14] Fan, L. T., Chen, S. J., and Watson, C. A. Solids Mixing. *Industrial and Engineering Chemistry*, 62(7):53–69, 1970.
- [15] Rollins, D. K., Faust, D. L., and Jabas, D. L. A superior approach to indices in determining mixture segregation. *Powder Technology*, 84:277–282, 1995.
- [16] Wilrich, P.-Th. In *Proceedings of VIIth International Workshop on Intelligent Statistical Quality Control September 5-7, 2001*. Institute for Improvement in Quality and Productivity, University of Waterloo, Waterloo, Canada., 2001.
- [17] SAS version 8.02. SAS Institute, Inc., Cary, NC., 1999.
- [18] PQRI. The use of stratified sampling of blend and dosage units to demonstrate adequacy of mix for powder blends. www.pqri.org, September 3, 2001. DRAFT.
- [19] Montgomery, D. C. *Design and Analysis of Experiments*. John Wiley & Sons, Inc., fifth edition edition, 2001.
- [20] Guide to Inspections of oral Solid Dosage forms Pre/Post Approval Issues for Development and Validation. FDA Division of Field Investigations, January 1994.
- [21] Kræmer, J., Svensson, J. R., and Melgaard, H. Sampling Bias in blending Validation and a Different Approach to Homogeneity Assessment. *Drug Development and Industrial Pharmacy*, 25(2), 1999.
- [22] Berman, J., Elinski, D. E., Gonzales, C. R., Hofer, J. D., Jimenez, P. J., Planchard, J. A., Tlachac, R. J., and Vogel, P. F. Technical Report No. 25

- Blend uniformity Analysis: Validation and in-Process Testing. Technical Report 25, PDA Solid Dosage Process Validation Committee, 1997.
- [23] Berman, J., Schoeneman, A. , and Shelton, J. T. Unit Dose Sampling: A Tale of Two thieves. *Drug Development an Industrial Pharmacy*, 22(11):1121–1132, 1996.
- [24] Bartlett, M. S. and Kendall, D. G. The Statistical Analysis of Variance-heterogeneity and the Logarithmic Transformation. *Journ. Roy. Statist. Soc. Suppl. VIII*, 1946.
- [25] McCullagh, P. and Nelder, J. A. *Generalized Linear Models*. Chapman and Hall, London, second edition edition, 1983.
- [26] U.S. Pharmacopeia, USP 24-NF19, Second Supplement, July, 2000.
- [27] Pharmacopeial Previews; <905> Uniformity of Dosage Units, 1997.
- [28] Harmonization, <905> Uniformity of Dosage Units. *Pharmacopeial Forum*, 27(3):2615 – 2619, 2001.
- [29] Members of the Statistics Working Group Pharmaceutical Research and Manufacturers of America (PhRMA). Uniformity - Alternative to the USP Pharmacopeial. *Pharmacopeial Forum*, 25(2):7939 – 7948, 1999.
- [30] S-PLUS. Insightful Corp., 1700 West Lake Avenue North, Suite 500, Seattle, WA 98109, 2000.
- [31] Melgaard, H. and Thyregod, P. Acceptance sampling by variables under measurement uncertainty. In Wilrich P.-Th. Lenz H.-J., editor, *Frontiers in Statistical Quality Control 5*, Heidelberg, 1997. Physica-Verlag.
- [32] Wilrich, P.-Th. Part 1: Sampling inspection, Single sampling plans for inspection by variables in the presence of measurement error. *Allgemeines Statistisches Archiv*, 84:239–250, 2000.
- [33] Lieberman and Resnikoff. Sampling plans for Inspection by Variables. *Journ. Amer. Statist. Assoc.*, (50):457–516, 1955.
- [34] ISO 2859: Sampling procedures and charts for inspection by attributes, Part 1 to 4. International Organization for Standardization, Geneva, 1999.

- [35] ISO 3951: Sampling procedures and charts for inspection by variables for percent nonconforming. International Organization for Standardization, Geneva, 1989.
- [36] Quality System Requirements, QS - 9000, 3rd edition. Automotive Industry Action Group, Soutfield, Michigan.
- [37] MIL-STD-1916 Department of Defense Test Method Standard - Department of Defense Preferred Methods for Acceptance of Product. US Department of Defense, 1996.
- [38] EEC Council Directive No. 75/106/EEC of 19. December 1974, pp. 1-13 . Technical report, Official Journal of the European Communities L42, 15. february 1975.
- [39] EEC Council Directive of 20. January 1976, pp. 1-11. Technical report, Official Journal of the European Communities L46, 21. February 1999.
- [40] ISO 11462-1: Guidelines for implementation of statistical process control - Part 1 Elements of SPC. International Organization for Standardization. Geneva, 2001.
- [41] ISO 10576-1: Statistical methods - Guidelines for the evaluation of conformity with specified requirements - Part 1: General principles. International Organization for Standardization. Geneva, Final Draft submitted for voting 2002.
- [42] < 905 > USP Uniformity of Dosage Units. “USP XXI”, Mack Printing Company, 1984.
- [43] Holst, E., Thyregod, P., and Wilrich, P-Th. On Conformity Testing and the use of Two-stage Procedures. *International Statistical Review*, 69:419–432, 2001.
- [44] *Mathematical Statistics, Basic Ideas and Selected Topics*, volume 1. Prentice-Hall, Upper Saddle River, 2 edition, 2001.
- [45] Johnson, N. L., Kotz, S., and Balakrishnan, N. *Continuous Univariate Distributions*, volume 1. Wiley, New York, second edition, 1994.

- [46] Hahn, G. J. and Meeker, W. Q. *Statistical Intervals, A Guide for Practitioners*. John Wiley & Sons, New York, 1991.
- [47] *Probability and Statistics for Engineers and Scientists*. Prentice-Hall, Upper Saddle River, 6 edition, 1998.
- [48] ISO 2854: Statistical interpretation of data - Techniques of estimation relating to means and variances. International Organization for Standardization. Geneva., 1976.
- [49] ISO 3494: Statistical interpretation of data - Power of tests relating to means and variances. International Organization for Standardization. Geneva., 1976.
- [50] Bergum, J. S. Constructing Acceptance Limits for Multiple Stage Tests. *Drug Development and Industrial Pharmacy*, 16:2153–2166, 1990.
- [51] Bray, D. F., Lyon, D. A., and Burr, I. W. Three Class Attributes Plans in Acceptance Sampling. *Technometrics*, 15, 1973.
- [52] International Commission on Microbiological Specifications for Foods (ICMSF) of the International Union of Microbiological Societies, University of Toronto Press, Toronto. *Sampling for microbiological analysis: principles and specific applications*, 2 edition, 1986.
- [53] Dahms, S. and Hildebrandt. Some remarks on the Design of Three-Class Sampling Plans. *Journal of Food Protection*, 61, 1998.
- [54] Wallis, W. A. *Lot quality measured by proportion defective. Chapter 1 in Selected Techniques of Statistical Analysis (Eisenhart, Hastay and Wallis, eds)*. McGraw-Hill Book Co. Inc., New York, 1947.
- [55] MIL-STD-414 Sampling procedures and Tables for Inspection by Variables for Percent Defective. Superintendent of Documents, United States Government Printing Office, Washington, D.C., 1957.
- [56] American National Standard: Sampling Procedures and Tables for Inspection by Variables for Percent Nonconformance, ANSI/ASQC Standard Z1.9. American National Standards Institute, 1980.

- [57] Schilling, E. G. Revised variables acceptance sampling standards - ANSI Z1.9 (1980) and ISO 3951 (1980). *Journal of Quality Technology*, 13:131–138.
- [58] Boulanger, M., Johnson, M., Perrouchet, C., and Thyregod, P. International statistical standards. Their place in the life-cycle of products and services. *International Statistical Review*, (67):151–173, 1999.
- [59] Bruhn-Suhr, M., Krumbholz, W., and Lenz, H.-J. On the Design of Exact Single Sampling Plans by Variables. In Wilrich P.-Th. Lenz H.-J., editor, *Frontiers in Statistical Quality Control*, volume 4, pages 83–90. Physica Verlag, Heidelberg, 1992.
- [60] Lei, D.-H. and Vardeman, S. B. The LRT method of constructing a two-sided “variables” acceptance region and its comparison with other methods. *Communications in Statistics, Part A – Theory and Methods*, (27):329–351, 1998.
- [61] Lehmann, E. L. *Testing Statistical Hypotheses*. Wiley, second edition edition, 1986.
- [62] PhRMA. Recommendations for a Globally Harmonized Uniformity of Dosage Units Test. *Pharmacoepial Forum*, 25(4):8609–8624, 1999.
- [63] Wilrich, P.-Th. Single sampling plans for inspection by variables under a variance component situation. In *Proceedings of VIIth International Workshop on Intelligent Statistical Quality Control September 5-7, 2001.*, Waterloo, Canada. Institute for Improvement in Quality and Productivity, University of Waterloo.
- [64] Yalkowsky, S.H. and Bolton, S. Particle Size and Content Uniformity. *Pharmaceutical Research*, 7:962–966, 1990.
- [65] Stoyan, D., Kendall, W.S, and Mecke, J. *Stochastic Geometry and its Applications*. Wiley, second edition, 1995.
- [66] Johnson, N.L., Kemp, A.W., and Kotz, S. *Univariate Discrete Distributions*. Wiley, New York, 2 edition, 1993.

Ph. D. theses from IMM

1. **Larsen, Rasmus.** (1994). *Estimation of visual motion in image sequences.* xiv + 143 pp.
2. **Rygaard, Jens Moberg.** (1994). *Design and optimization of flexible manufacturing systems.* xiii + 232 pp.
3. **Lassen, Niels Christian Krieger.** (1994). *Automated determination of crystal orientations from electron backscattering patterns.* xv + 136 pp.
4. **Melgaard, Henrik.** (1994). *Identification of physical models.* xvii + 246 pp.
5. **Wang, Chunyan.** (1994). *Stochastic differential equations and a biological system.* xxii + 153 pp.
6. **Nielsen, Allan Aasbjerg.** (1994). *Analysis of regularly and irregularly sampled spatial, multivariate, and multi-temporal data.* xxiv + 213 pp.
7. **Ersbøll, Annette Kjær.** (1994). *On the spatial and temporal correlations in experimentation with agricultural applications.* xviii + 345 pp.
8. **Møller, Dorte.** (1994). *Methods for analysis and design of heterogeneous telecommunication networks.* Volume 1-2, xxxviii + 282 pp., 283-569 pp.
9. **Jensen, Jens Christian.** (1995). *Teoretiske og eksperimentelle dynamiske undersøgelser af jernbanekøretøjer.* viii + 174 pp.
10. **Kuhlmann, Lionel.** (1995). *On automatic visual inspection of reflective surfaces.* Volume 1, xviii + 220 pp., (Volume 2, vi + 54 pp., fortrolig).
11. **Lazarides, Nikolaos.** (1995). *Nonlinearity in superconductivity and Josephson Junctions.* iv + 154 pp.
12. **Rostgaard, Morten.** (1995). *Modelling, estimation and control of fast sampled dynamical systems.* xiv + 348 pp.
13. **Schultz, Nette.** (1995). *Segmentation and classification of biological objects.* xiv + 194 pp.

14. **Jørgensen, Michael Finn.** (1995). *Nonlinear Hamiltonian systems*. xiv + 120 pp.
15. **Balle, Susanne M.** (1995). *Distributed-memory matrix computations*. iii + 101 pp.
16. **Kohl, Niklas.** (1995). *Exact methods for time constrained routing and related scheduling problems*. xviii + 234 pp.
17. **Rogon, Thomas.** (1995). *Porous media: Analysis, reconstruction and percolation*. xiv + 165 pp.
18. **Andersen, Allan Theodor.** (1995). *Modelling of packet traffic with matrix analytic methods*. xvi + 242 pp.
19. **Hesthaven, Jan.** (1995). *Numerical studies of unsteady coherent structures and transport in two-dimensional flows*. Risø-R-835(EN) 203 pp.
20. **Slivsgaard, Eva Charlotte.** (1995). *On the interaction between wheels and rails in railway dynamics*. viii + 196 pp.
21. **Hartelius, Karsten.** (1996). *Analysis of irregularly distributed points*. xvi + 260 pp.
22. **Hansen, Anca Daniela.** (1996). *Predictive control and identification - Applications to steering dynamics*. xviii + 307 pp.
23. **Sadegh, Payman.** (1996). *Experiment design and optimization in complex systems*. xiv + 162 pp.
24. **Skands, Ulrik.** (1996). *Quantitative methods for the analysis of electron microscope images*. xvi + 198 pp.
25. **Bro-Nielsen, Morten.** (1996). *Medical image registration and surgery simulation*. xxvii + 274 pp.
26. **Bendtsen, Claus.** (1996). *Parallel numerical algorithms for the solution of systems of ordinary differential equations*. viii + 79 pp.
27. **Lauritsen, Morten Bach.** (1997). *Delta-domain predictive control and identification for control*. xxii + 292 pp.

28. **Bischoff, Svend.** (1997). *Modelling colliding-pulse mode-locked semiconductor lasers.* xxii + 217 pp.
29. **Arnbjerg-Nielsen, Karsten.** (1997). *Statistical analysis of urban hydrology with special emphasis on rainfall modelling.* Institut for Miljøteknik, DTU. xiv + 161 pp.
30. **Jacobsen, Judith L.** (1997). *Dynamic modelling of processes in rivers affected by precipitation runoff.* xix + 213 pp.
31. **Sommer, Helle Mølgaard.** (1997). *Variability in microbiological degradation experiments - Analysis and case study.* xiv + 211 pp.
32. **Ma, Xin.** (1997). *Adaptive extremum control and wind turbine control.* xix + 293 pp.
33. **Rasmussen, Kim Ørskov.** (1997). *Nonlinear and stochastic dynamics of coherent structures.* x + 215 pp.
34. **Hansen, Lars Henrik.** (1997). *Stochastic modelling of central heating systems.* xxii + 301 pp.
35. **Jørgensen, Claus.** (1997). *Driftoptimering på kraftvarmesystemer.* 290 pp.
36. **Stauning, Ole.** (1997). *Automatic validation of numerical solutions.* viii + 116 pp.
37. **Pedersen, Morten With.** (1997). *Optimization of recurrent neural networks for time series modeling.* x + 322 pp.
38. **Thorsen, Rune.** (1997). *Restoration of hand function in tetraplegics using myoelectrically controlled functional electrical stimulation of the controlling muscle.* x + 154 pp. + Appendix.
39. **Rosholm, Anders.** (1997). *Statistical methods for segmentation and classification of images.* xvi + 183 pp.
40. **Petersen, Kim Tilgaard.** (1997). *Estimation of speech quality in telecommunication systems.* x + 259 pp.

41. **Jensen, Carsten Nordstrøm.** (1997). *Nonlinear systems with discrete and continuous elements.* 195 pp.
42. **Hansen, Peter S.K.** (1997). *Signal subspace methods for speech enhancement.* x + 226 pp.
43. **Nielsen, Ole Møller.** (1998). *Wavelets in scientific computing.* xiv + 232 pp.
44. **Kjems, Ulrik.** (1998). *Bayesian signal processing and interpretation of brain scans.* iv + 129 pp.
45. **Hansen, Michael Pilegaard.** (1998). *Metaheuristics for multiple objective combinatorial optimization.* x + 163 pp.
46. **Riis, Søren Kamaric.** (1998). *Hidden markov models and neural networks for speech recognition.* x + 223 pp.
47. **Mørch, Niels Jacob Sand.** (1998). *A multivariate approach to functional neuro modeling.* xvi + 147 pp.
48. **Frydendal, Ib.** (1998.) *Quality inspection of sugar beets using vision.* iv + 97 pp. + app.
49. **Lundin, Lars Kristian.** (1998). *Parallel computation of rotating flows.* viii + 106 pp.
50. **Borges, Pedro.** (1998). *Multicriteria planning and optimization. - Heuristic approaches.* xiv + 219 pp.
51. **Nielsen, Jakob Birkedal.** (1998). *New developments in the theory of wheel/rail contact mechanics.* xviii + 223 pp.
52. **Fog, Torben.** (1998). *Condition monitoring and fault diagnosis in marine diesel engines.* xii + 178 pp.
53. **Knudsen, Ole.** (1998). *Industrial vision.* xii + 129 pp.
54. **Andersen, Jens Strodl.** (1998). *Statistical analysis of biotests. - Applied to complex polluted samples.* xx + 207 pp.

55. **Philipsen, Peter Alshede.** (1998). *Reconstruction and restoration of PET images.* vi + 132 pp.
56. **Thygesen, Uffe Høgsbro.** (1998). *Robust performance and dissipation of stochastic control systems.* 185 pp.
57. **Hintz-Madsen, Mads.** (1998). *A probabilistic framework for classification of dermatoscopic images.* xi + 153 pp.
58. **Schramm-Nielsen, Karina.** (1998). *Environmental reference materials methods and case studies.* xxvi + 261 pp.
59. **Skyggebjerg, Ole.** (1999). *Acquisition and analysis of complex dynamic intra- and intercellular signaling events.* 83 pp.
60. **Jensen, Kåre Jean.** (1999). *Signal processing for distribution network monitoring.* xv + 199 pp.
61. **Folm-Hansen, Jørgen.** (1999). *On chromatic and geometrical calibration.* xiv + 238 pp.
62. **Larsen, Jesper.** (1999). *Parallelization of the vehicle routing problem with time windows.* xx + 266 pp.
63. **Clausen, Carl Balslev.** (1999). *Spatial solitons in quasi-phase matched structures.* vi + (flere pag.)
64. **Kvist, Trine.** (1999). *Statistical modelling of fish stocks.* xiv + 173 pp.
65. **Andresen, Per Rønsholt.** (1999). *Surface-bounded growth modeling applied to human mandibles.* xxii + 125 pp.
66. **Sørensen, Per Settergren.** (1999). *Spatial distribution maps for benthic communities.*
67. **Andersen, Helle.** (1999). *Statistical models for standardized toxicity studies.* viii + (flere pag.)
68. **Andersen, Lars Nonboe.** (1999). *Signal processing in the dolphin sonar system.* xii + 214 pp.

69. **Bechmann, Henrik.** (1999). *Modelling of wastewater systems*. xviii + 161 pp.
70. **Nielsen, Henrik Aalborg.** (1999). *Parametric and non-parametric system modelling*. xviii + 209 pp.
71. **Gramkow, Claus.** (1999). *2D and 3D object measurement for control and quality assurance in the industry*. xxvi + 236 pp.
72. **Nielsen, Jan Nygaard.** (1999). *Stochastic modelling of dynamic systems*. xvi + 225 pp.
73. **Larsen, Allan.** (2000). *The dynamic vehicle routing problem*. xvi + 185 pp.
74. **Halkjær, Søren.** (2000). *Elastic wave propagation in anisotropic inhomogeneous materials*. xiv + 133 pp.
75. **Larsen, Theis Leth.** (2000). *Phosphorus diffusion in float zone silicon crystal growth*. viii + 119 pp.
76. **Dirscherl, Kai.** (2000). *Online correction of scanning probe microscopes with pixel accuracy*. 146 pp.
77. **Fisker, Rune.** (2000). *Making deformable template models operational*. xx + 217 pp.
78. **Hultberg, Tim Helge.** (2000). *Topics in computational linear optimization*. xiv + 194 pp.
79. **Andersen, Klaus Kaae.** (2001). *Stochastic modelling of energy systems*. xiv + 191 pp.
80. **Thyregod, Peter.** (2001). *Modelling and monitoring in injection molding*. xvi + 132 pp.
81. **Schjødt-Eriksen, Jens.** (2001). *Arresting of collapse in inhomogeneous and ultrafast Kerr media*.
82. **Bennetsen, Jens Christian.** (2001). *Numerical simulation of turbulent airflow in livestock buildings*. xi + 205 pp + Appendix.

83. **Højen-Sørensen, Pedro A.d.F.R.** (2001). *Approximating methods for intractable probabilistic models: - Applications in neuroscience.* xi + 104 pp + Appendix.
84. **Nielsen, Torben Skov.** (2001). *On-line prediction and control in non-linear stochastic systems.* xviii + 242 pp.
85. **Öjelund, Henrik.** (2001). *Multivariate calibration of chemical sensors.* xviii + 182 pp.
86. **Adeler, Pernille Thorup.** (2001). *Hemodynamic simulation of the heart using a 2D model and MR data.* xv + 179 pp.
87. **Nielsen, Finn Årup.** (2001). *Neuroinformatics in functional neuroimaging.* 330 pp.
88. **Kidmose, Preben.** (2001). *Blind separation of heavy tail signals.* viii + 136 pp.
89. **Hilger, Klaus Baggesen.** (2001). *Exploratory analysis of multivariate data.* xxiv + 180 pp.
90. **Antonov, Anton.** (2001). *Object-oriented framework for large scale air pollution models.* 156 pp. + (flere pag).
91. **Poulsen, Mikael Zebbelin.** (2001). *Structural analysis of DAEs.* 192 pp.
92. **Keijzer, Maarten.** (2001). *Scientific discovery using genetic programming.*
93. **Sidaros, Karam.** (2002). *Slice profile effects in MR perfusion imaging using pulsed arterial spin labelling.* xi + 202 pp.
94. **Kolenda, Thomas.** (2002). *Adaptive tools in virtual environments. - Independent component analysis for multimedia.* xiii + 117 pp.
95. **Berglund, Ann-Charlotte.** (2002). *Nonlinear regularization with applications in geophysics.* xiii + 116 pp.
96. **Pedersen, Lars.** (2002). *Analysis of two-dimensional electrophoresis Gel images.* xxii + 161 pp.

97. **Rasmussen, Thomas Marthedal.** (2002). *Interval logic. Proof theory and theorem proving.* xii + 150 pp.
98. **Vistisen, Dorte.** (2002). *Models and methods for hot spot safety work.* ix + 168 pp.
99. **Madsen, Camilla.** (2002). *Statistical methods for assessment of blend homogeneity.* xviii + 246 pp.
100. **Holten-Lund, Hans.** (2002). *Design for scalability in 3D computer graphics architectures.* x + 169 pp.
101. **Bache, Morten.** (2002). *Quantum noise and spatio-temporal pattern-formation in nonlinear optics.* ix + 123 pp.
102. **Stidsen, Thomas Jacob Kjær.** (2002). *Optimisation problems in optical network design.* xii + 55 + Appendix.
103. **Andersen, Irene Klærke.** (2002). *Dynamic contrast enhanced MRI for perfusion quantification.* iv + 181 pp.
104. **Xia, Fujie.** (2002). *The dynamics of the three-piece-freight truck.* x + 205 pp.
105. **Johansen, Steffen Kjær.** (2002). *All-optical signal processing in quadratic nonlinear materials.* x + 90 pp.
106. **Nyeland, Martin Erik.** (2002). *Sampling strategy and statistical modelling of exposure.* viii + 39 + appendix.