Citation:

Böhning, D. & Dammann, U.-P. (2003). The application of methods of meta-analysis for heterogeneity modelling in quality control and assurance. In R. Schulze, H. Holling & D. Böhning (Eds.), *Meta-analysis: New developments and applications in medical and social sciences* (pp. 155-164). Hogrefe & Huber.

10

The Application of Methods of Meta-Analysis for Heterogeneity Modeling in Quality Control and Assurance

Dankmar Böhning
Working Group: Biometry and Epidemiology
Institute for International Health, Joint Center for Humanities and Health Sciences
Humboldt-University in Berlin / Free University Berlin

Uwe-Peter Dammann
Asta-Medica, Werk Künsebeck
Halle-Künsebeck

Summary

In the past few years meta-analysis has become increasingly popular in many areas of science such as medicine and pharmacy, psychology, and other social sciences. In these areas of application meta-analyses have been performed in order to obtain a pooled estimate of various single studies. Obtaining a single summary measure implicitly assumes homogeneity of these studies, that is, the results of individual studies differ only by chance. In this case a combined estimate of the individual studies provides a powerful and important result. However, this pooled estimate may be seriously misleading if study conditions are heterogenous. Thus, an approach which considers meta-analysis as a study over studies has increasingly been advocated. This approach seeks to investigate heterogeneity between studies. An important feature of this type of meta-analysis lies in the fact that it tries to identify factors which cause heterogeneity. It is the aim of this contribution, in corporation with the unit of quality assurance of ASTA Medica at location Künsebeck, to extend this approach appropriately to the area of quality control, where batches of the produced

goods replace the role of studies in medicine or the social sciences. Clearly, in this setting an investigation of heterogeneity is equally attractive, since identification and modeling of heterogeneity help to improve the production process.

10.1 INTRODUCTION AND PREVIEW

The chapter reviews an approach which enables a global perspective on aspects of homogeneity and heterogeneity which occurs in quality control and quality assurance in the pharmaceutical industry. In conventional meta-analysis, investigations are done in such a way that a specific measure can be computed utilizing numerous single studies. Frequently, statistical questions of efficiency are dominating in the literature (Hedges & Olkin, 1985). Efficiency is achieved by pooling the various single studies, thus yielding an increased sample size. This procedure, no doubt, is of great benefit, if the various studies to be combined in the meta-analysis have emerged under comparable conditions and are different in a statistical sense only by chance. This is the situation of *homogeneity*. However, pooled analysis is often considered problematic if study conditions are heterogenous, especially if the interpretation of pooled estimators is kept in a traditional way.

The chapter at hand underlines parallel aspects of meta-analysis and quality control. The cornerstone of this analogy are the numerous batches which are drawn in quality control for monitoring purposes, which play the role of the single studies in meta-analysis. Here, measures of interest are frequently count variables (counts of contamination particles) or other quality indices. In this situation – even if homogeneity conditions are present – deviations from a given standard might occur as well. It is quite important whether these deviations might have emerged from a homogenous process (as random variations) or are due to certain *heterogeneities* present in the production process. By means of the mixture distribution analysis, we are able to model potentially present heterogeneity and, further on, to classify each batch into one of the heterogeneity components. This might allow the researcher to diagnose certain common attributes and therefore enables him to explore the causes of heterogeneity.

10.2 LEGAL BACKGROUND FOR PHARMACEUTICAL PRODUCTION

Pharmaceutical production of drug products and drug substances is regulated worldwide by the rules of Good Manufacturing Practices. For Europe and Germany, producers have to follow the regulations of

- Arzneimittelgesetz (AMG)
- EU-Guideline for Good Manufacturing Practices (1989)

- "Betriebsverordnung für pharmazeutische Unternehmer" (PharmBetrV 1994)

Production and quality control of drug products and drug substances have to recognize state of the art and current worldwide practices in accordance with the application. All procedures used in production and quality control must be validated and regularly revalidated. Drug products are mainly produced in batches, which should conform with the specification from batch to batch. Drug products brought into the market should be produced and controlled according to the application and the quality has to be confirmed before a batch can be released for distribution.

The quality of a drug product or a drug substance is defined by identity, assay, chemical, physical and biological properties. A batch is the quantity of a drug produced under suitable uniform conditions to guarantee a homogeneous quality.

10.3 THE TASKS AND OBJECTIVES OF QUALITY ASSURANCE IN PHARMACEUTICAL INDUSTRY

The production of drugs is accompanied by

- batch- and product related in-process controls (on-line)
- batch- and product related controls (off-line)
- not batch and not product related controls

Parenteral drugs are products which have to comply with additional, specific properties like sterility and are essentially free of visible particles because of their parenteral application. Sterility is controlled by a sterility test which is a destructive test on limited samples of a batch. In connection with in-process controls for the clean environment of rooms, air, surface, and personnel hygiene during production, especially parenteral drugs produced by aseptic processing sterility can be assured in all parts of a batch.

Each parenteral container is controlled by a 100%-inspection for particulate matter. The quality of this inspection is controlled by samples which are again inspected for subvisual particles. These are destructive tests on a limited number of samples. The quality is evaluated on the basis of a quality index like the one which can be found in the Deutscher Arzneimittel Codex (DAC), Codex Probe no. 5. The particulate matter is evaluated for particles which can be seen easily, well, or with difficulties.

For instance:

- No visible particle: no point
- Particle difficult to be seen within 5 seconds: one point
- Particle easily to be seen within 5 seconds: two points
- Particles to be seen immediately and in higher numbers: ten points

The formula for evaluation is: $Q_{TR} = \frac{A}{N}$, where A stands for the number of points recorded by three test persons and N stands for the number of controlled containers.

The results of all controls for one batch and from batch to batch is very important for the evaluation of the quality and the release for distribution. Trends for a homogeneous or heterogeneous process should be addressed and recognized as soon as it happens. Statistical evaluation of all available data is essential for the routine evaluation of the drug quality.

10.4 META-ANALYTIC MODELING OF DATA OCCURRING IN QUALITY ASSURANCE

Very often quality assurance is based on the availability of a number of batches each having a certain number of items. For example, we might consider again Q_{TR} and define X as

$$X = \text{Number of times with } Q_{TR} \text{ positive in a series of } n \text{ investigations.}$$

This is best demonstrated by means of an example which is taken from the book of Derman and Ross (1997). The data are provided in Table 10.1 and visualized by means of a confidence interval plot (proportion with 95% confidence interval) in Figure 10.1.

Table 10.1 Number of Defective Items for 20 Batches of 200 Items Each

Batch	Number of Defectives	Batch	Number of Defectives
1	24	11	4
2	22	12	13
3	12	13	17
4	13	14	5
5	15	15	9
6	11	16	0
7	25	17	19
8	16	18	0
9	23	19	22
10	14	20	17

As has been pointed out in the literature (Petitti, 1994), the area of meta-analysis has received various impulses during its historic development. In psychology, the development of measures was achieved which could be suitably used for meta-analysis such as the standardized effect difference. Another impulse was the development of suitable statistical methods such as the appropriate form of a *pooled mean*. Meta-analysis experienced tremendous impulses by means of embedding important application areas such as evaluation research or health reporting. It is hoped that both areas discussed in this chapter,

namely quality control and assurance and meta-analysis, experience a similar impulse from each other.

It is quite obvious that in quality control the single batch can play the role of a single study in conventional meta-analysis. This can avoid various techniques including control charts and repeated testing, which can be statistically flawed. For example, if 20 binomial tests are employed for the data provided in Table 10.1, it can be expected that one of these will show a significant deviation from a desired standard though there is in fact no deviation from the desired standard (process is still in control). Similarly, if control charts are used, it is well-known that the boundaries of these charts are reached for some batch, though the process is still in control. As a consequence, investigators in quality assurance are forced to investigate for a non-existing source of deviation of the production process.

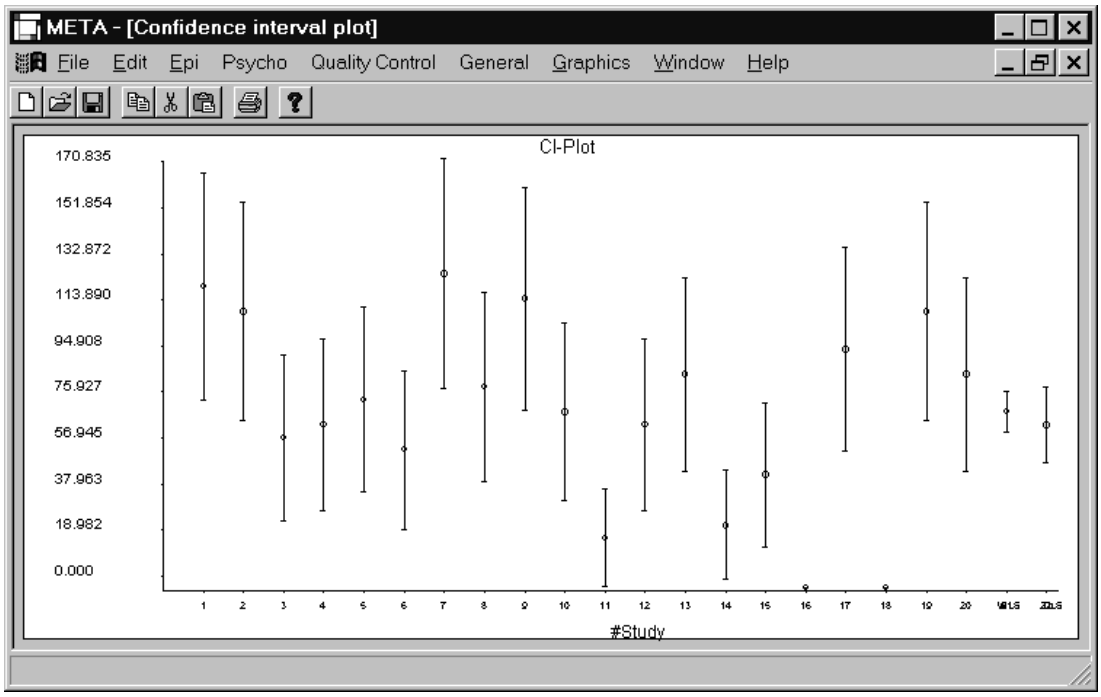

Figure 10.1 Confidence interval plot from the package META for a textbook example of proportion of defective items for 20 batches with 200 items each.

10.5 THE PROBLEM OF HETEROGENEITY

In fact, we are interested in separating *random deviations*, which are occurring always in non-deterministic systems[1], and *systematic deviations*. Only the latter are relevant and prone for further investigation and research.

How can we accomplish this separation? The first step is to model the situation when the process is in control, which is called the situation of *homogeneity*. Typically, it is possible to derive some probability distribution for the measure of interest under homogeneity. We call the associated density of the measure of interest X: $f(x,\theta)$, where θ is some parameter involved in this density. In our example, the *number of defective items*, X, follows a binomial distribution with density $f(x,\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$, where n is the size of the batch and the parameter θ corresponds to the allowed number of defectives.

The question at hand is: What will happen if a deviation (loss in quality) occurs and how is this reflected in the statistical model? Clearly, if this happens, homogeneity conditions no longer hold and the simple statistical model $f(x,\theta)$ will no longer be correct.

There are some simple tests available which allow to diagnose this situation rather quickly. One of these tests is based upon the defined as

$$\chi^2 = \sum_{i=1}^{k} \frac{(X_i - E(X_i))^2}{\mathrm{Var}(X_i)}.$$

Typically, $E(X_i)$ and $\mathrm{Var}(X_i)$ will be functions of the unknown parameter θ and plug-in-estimates must be utilized. These plug-in estimators must be constructed with care to achieve χ^2-distribution under homogeneity, at least approximately. To give a demonstration, we note that in our binomial quality control example $E(X_i) = n\theta_i$ and $\mathrm{Var}(X_i) = n\theta_i(1-\theta_i)$, which might lead to the plug-in estimates $\widehat{E(X_i)} = X_i$ and $\widehat{\mathrm{Var}(X_i)} = X_i(1 - X_i/n)$. It can be shown that the associated distribution under homogeneity is quite different from a χ^2-distribution with $k-1$ degrees of freedom if sample sizes per batch, n, are small or moderate, even if the number of batches k becomes large. The right thing to do here turns out to be a variance estimate utilizing information from all batches: $\widehat{\mathrm{Var}(X_i)} = S_k(1 - S_k/n)$, where $S_k = \sum_{i=1}^{k} X_i/k$. The associated χ^2-statistic (with $E(X_i) = S_k$) can be shown to be validly approximated by a χ^2-distribution with $k-1$ degrees of freedom even for small batch size n (like $n = 5$). For further discussion, see Böhning (2000b) as well as Hartung and Knapp (Chapter 4, this volume). To finish this aspect, we find a value of $\chi^2 = 70.41$ with 19 degrees of freedom for the data of Table 10.1, which indicates strongly the presence of heterogeneity.

[1]The question which system is deterministic and which is not is a mere philosophical question. Our point of view is that it is appropriate and useful to consider stochastic variation even when measurements and processes are done with the highest accuracy.

In the following section we will concentrate on the aspect: What can be done if heterogeneity is present?

10.6 MODELING HETEROGENEITY USING MIXTURE DISTRIBUTIONS

If heterogeneity is present it is implied that the proportion of defectives in the batch is deviating in a systematic way from the required standard, in other words, it can be assumed that the hypothesis $\theta_1 = \theta_2 = \ldots = \theta_k = \theta$ is wrong and it is more reasonable to assume that for certain parts of the population of all possible batches a value (for the proportion of defectives) of θ_1 – for other parts a value of θ_2 – is valid and so forth. That is, the population of possible batches consists of a proportion p_j of batches with θ_j, for $j = 1, \ldots, k$. It can be shown (Böhning, 2000a) that in this situation X_i has a mixture distribution

$$f(x_i, P) = \sum_{j=1}^{k} f(x_i, \theta_j) p_j$$

which takes the form of a mixture of binomial distributions for our textbook example:

$$f(x_i, P) = \sum_{j=1}^{k} \binom{n}{x_i} \theta_j^{x_i} (1 - \theta_j)^{n - x_i} p_j. \tag{10.1}$$

The distribution which gives probability mass p_j to θ_j is called *mixing distribution* and is denoted by P. To estimate the parameters involved in Equation 10.1, in other words the mixing distribution P, we use maximum likelihood estimation including the number of components in the mixture k. This can be accomplished with the computer package `C.A.MAN` (see Böhning, Schlattmann, & Lindsay, 1992; Böhning, Dietz, & Schlattmann, 1998). The associated maximum likelihood estimate of k and θ_j, p_j for $j = 1, \ldots, k$ is called *nonparametric maximum likelihood estimate* (NPMLE) of the mixing distribution P. It is usually advisable to check whether the number of components k can be reduced, which can be accomplished by comparing log-likelihoods for reduced values of k such as $k - 1, k - 2, \ldots$ until no significant drop in the log-likelihood is notable. For these fixed values of k estimation is done via the EM-algorithm (Dempster, Laird, & Rubin, 1977).

To provide a demonstration for this technique, we study the data of Table 10.1 again and use the mixture model provided in Equation 10.1. Table 10.2 provides the results. There is empirical evidence for *heterogeneity* and that this heterogeneity consists of 3 components.

It can be seen that the population of batches can be separated into *three* components. One component consists of batches which are free of defective items (9.9%). The second component has 2.87 defective items per 100 (13.3%), whereas the last one has 8.6 defective items per 100, representing the majority of all batches (76.8%).

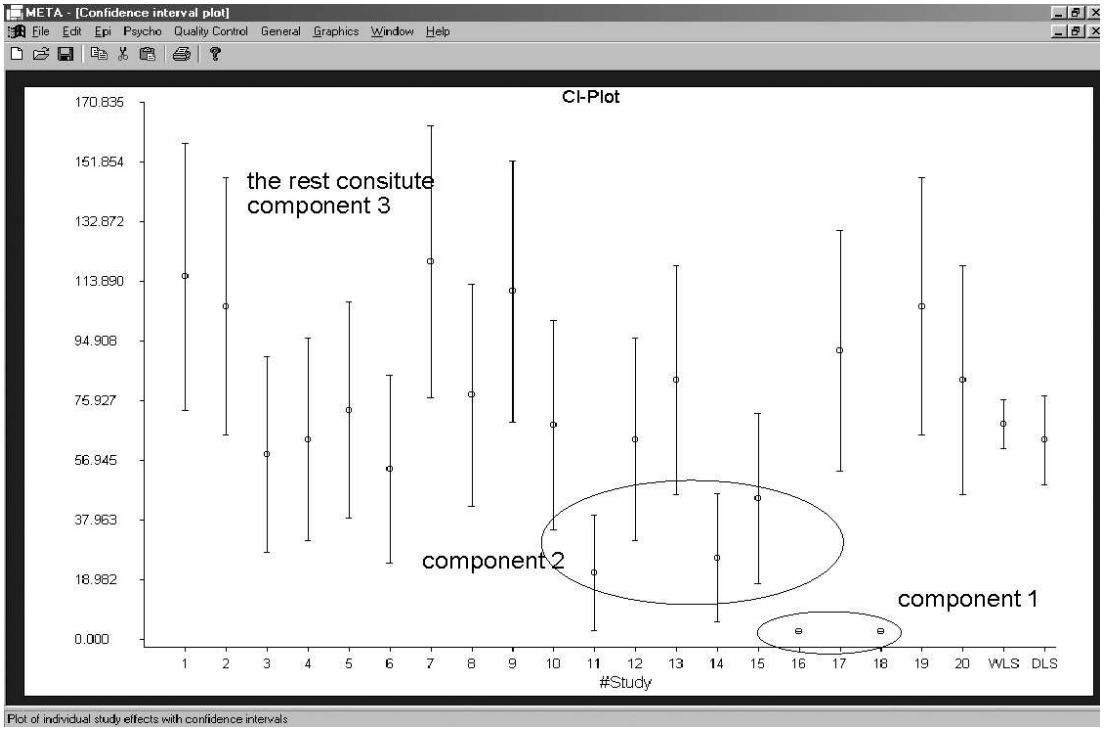

Figure 10.2 Classification of the batches into their associated components for the textbook example of proportion of defective items for 20 batches with 200 items each.

Table 10.2 Identification of Heterogeneity Structure for 20 Batches of 200 Items Each

Number of Components k	Log-Likelihood
4 (NPMLE)	−63.1454
3	−64.0984
2	−70.9835

Estimated Mixing Distribution for $k = 3$

Proportion θ_j	Weight p_j
0.0000	0.0996
0.0287	0.1326
0.0865	0.7678

Finally, it is even possible to allocate each observed (investigated) batch to one of the components in the mixture. This can be accomplished by utilizing Bayes theorem and calculate the posterior distribution of θ as

$$f(\theta_j|x_i) = \frac{f(x_i, \hat{\theta}_j)\hat{p}_j}{\sum_{l=1}^{k} f(x_i, \hat{\theta}_l)\hat{p}_l},$$

where $\hat{\theta}_j$ and $\hat{p}_j$ correspond to the maximum likelihood estimates identified in the previous estimation process. Each batch i with number of defectives X_i is allocated to that component j for which $f(\theta_j|x_i)$ is largest of all $j = 1, \ldots, k$. This is done for the data in Table 10.1 and the results are provided in Table 10.3. Figure 10.2 also visualizes this reclassification. This technique might enable the practitioner to search for common sources for the occurred heterogeneity and finally identify sources for the loss in quality standards.

Table 10.3 Classification of Each Batch Into the Components

Batch i	X_i	Component j	Batch i	X_i	Component j
1	24	3	11	4	2
2	22	3	12	13	3
3	12	3	13	17	3
4	13	3	14	5	2
5	15	3	15	9	2
6	11	3	16	0	1
7	25	3	17	19	3
8	16	3	18	0	1
9	23	3	19	22	3
10	14	3	20	17	3

10.7 DISCUSSION

We touched upon an approach which explicitly allows the modeling of heterogeneity. To do this, it is important to emphasize that an appropriate measure of interest (describing the quality standards) has to be chosen. Given the chosen measure of interest, it is furthermore equally important to find the corresponding statistical model under homogeneity conditions and further the associated mixture model which models potential heterogeneity. A variety of situations have been assembled to form a package `META` which allows the user in a simple way to analyze heterogeneity problems in his/her application. For details, see Schlattmann, Malzahn, and Böhning (Chapter 16, this volume).

REFERENCES

Böhning, D. (2000a). *Computer-assisted analysis of mixtures and applications. Meta-analysis, disease mapping, and others.* Boca Raton, FL: Chapman & Hall/CRC.

Böhning, D. (2000b). *The flaw in χ^2-heterogeneity tests: A revisit of the Neyman-Scott problem?* Unpublished report, Free University Berlin.

Böhning, D., Dietz, E., & Schlattmann, P. (1998). Recent developments in computer-assisted analysis of mixtures. *Biometrics*, *54*, 525–536.

Böhning, D., Schlattmann, P., & Lindsay, B. G. (1992). Computer-assisted analysis of mixtures (C.A.MAN): Statistical algorithms. *Biometrics*, *48*, 283–303.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B*, *39*, 1–38.

Derman, C., & Ross, S. M. (1997). *Statistical aspects of quality control.* San Diego, CA: Academic Press.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Orlando, FL: Academic Press.

Petitti, D. B. (1994). *Meta-analysis, decision analysis and cost-effectiveness analysis. Methods for quantitative synthesis in medicine.* Oxford: Oxford University Press.

Acknowledgments

This research was done under support of the *Bundesministerium für Bildung und Wissenschaft, Forschung und Technologie (BMBF).*