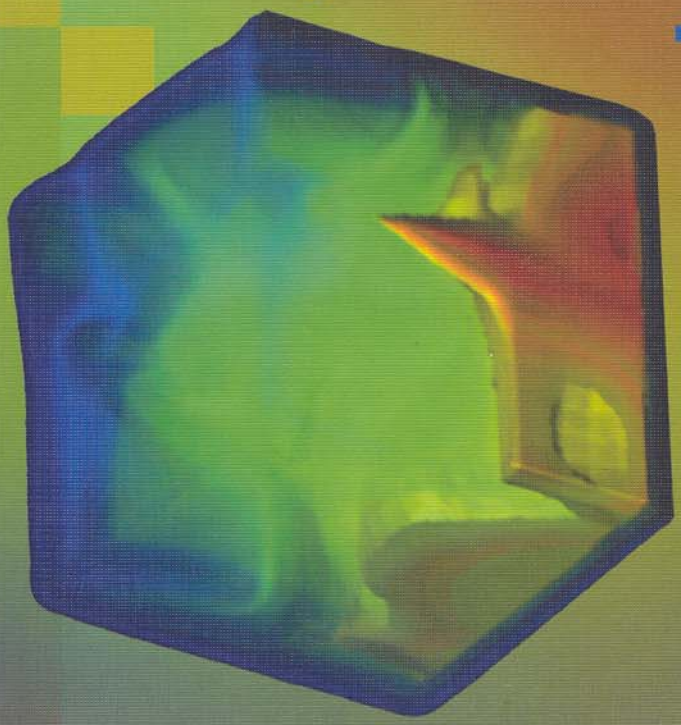
Willi Jäger
Hans-Joachim Krebs
Editors

MATHEMATICS
Key Technology for the Future

Joint Projects between Universities and Industry



Förderschwerpunkt
Mathematische Verfahren zur Lösung von Problemstellungen in Industrie und Wirtschaft

Springer

Table of Contents

The Application of Statistical Methods of Meta-Analysis for Heterogeneity Modelling in Medicine and Pharmacy, Psychology, Quality Control and Assurance

Dankmar Böhning[1], Uwe Malzahn[1], Peter Schlattmannn[1], Uwe-Peter Dammann[2], Wolfgang Mehnert[3], Heinz Holling[4], and Ralf Schulze[4]

[1] Department of Epidemiology, Free University Berlin, Haus 562, Fabeckstr. 60–62, 14195 Berlin, Germany, E-mail: boehning@zedat.fu-berlin.de
[2] Asta-Medica, Werk Künsebeck, Kantstr. 2, 33790 Halle-Künsebeck, Germany, E-mail: Uwe-Peter.Dammann@astamedica.de
[3] Department of Pharmaceutical Technology, Free University Berlin, Kelchstr. 31, 12169 Berlin, Germany, E-mail: mehnert@zedat.fu-berlin.de
[4] Occupational Psychology, University of Münster, Fliednerstr. 21, 48149 Münster, Germany, E-mail: holling@psy.uni-muenster.de

Abstract. 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, i.e. 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, increasingly an approach has been advocated which considers meta-analysis as a study over studies. 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 has been the focus on this project (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 modelling of heterogeneity helps to improve the production process.

1 Introduction

The paper reviews an approach which enables a global perspective on aspects of homogeneity and heterogeneity which occurrs when applying methods of meta-analysis to clinical studies in medicine and pharmacy, psychology and other social sciences, but also 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 and Olkin, [1985]), which is achieved by pooling the various single studies and, thus by, achieving 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 are kept in a traditional way.

1.1 Meta-Analysis of Clinical Studies in Medicine and Pharmacy

In clinical trials often a treatment group is compared with a control group, and the risk of some event (like survial after treatment) is compared between both groups. Let p_1 be the risk (probability) in the treatment group and p_0 the risk in the control group, then typical measures considered are the *relative risk* $RR = p_1/p_0$ or the risk difference $RD = p_1 - p_0$. These measures are estimated in several, say n, studies and then pooled in a summary measure, for example in the case of the risk difference we have $\sum_{i=1}^{n} w_i \hat{RD}_i$ where the weights w_i are proportional to the inverse variance $Var(\hat{RD}_i), i = 1, ..., n$. There are numerous examples of this kind of meta-analysis and a recent reference to introductory reviews is Normand ([1999]) or earlier Jones ([1995]).

1.2 Meta-Analysis of Experimental Studies in Psychology

In the social sciences, primarily in psychology, an effect measurure is computed for experimental or quasi-experimental studies which is often the standardized difference (difference of the means in treatment and control, then devided by the common standard deviation) or the correlation coefficient. A detailed discussion on the standardized difference is provided in Hedges and Olkin ([1985]), Cooper and Hedges ([1994]), and, in terms of the distributional aspects involved, in Malzahn, Böhning and Holling ([2000]). Details on the measure of the correlation coefficient are found in Hedges and Olkin ([1985]) and Cooper and Hedges ([1994]). Typically, the correlation coefficient ρ is used in it's Fisher-transformed version having an approximate variance of $Var\{log(z_i)\} = \frac{1}{m_i - 3}$, where $z_i = 0.5 \frac{1+\hat{\rho}_i}{1-\hat{\rho}_i}$ is the Fisher-transformation and m_i is the sample size of study i, $i = 1, ...n$. Consequently, the summary Fisher-transformed correlation coefficient takes on the simple form $\frac{\sum_{i=1}^{n} m_i z_i}{\sum_{i=1}^{n} m_i}$ which is popular for it's simplicity.

2 Meta-Analysis in Quality Control

The project and consequently the paper at hand investigates parallel aspects of meta-analyis 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. Measures of interest are here frequently count variables (counts of contamination particles) or other quality indeces. In this situation – even if homogeneity conditions are present – deviations from a given standard might occurr, 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 furtheron, to classify each batch into one of the heterogneity components. This might allow to diagnose certain common attributes and therefore be able to explore for the causes of heterogeneity.

2.1 Legal Background for Pharmaceutical Production

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

1. Arzneimittelgesetz (AMG)
2. EU-Guideline for Good Manufacturing Practices (1989)
3. "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 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.

2.2 The Tasks and Objectives of Quality Assurance in Pharmaceutical Industry

The production of drugs is accompanied by

1. batch-and product related in-process controls (on line)
2. batch- and product related controls (off line)
3. not batch and not product related controls

Parenteral drugs are products which have to comply with additional, specific properties like sterility and essentially free of visible particles because of their parenteral application. Sterility is controlled by a sterility test which is 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 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, good or difficult. For instance:

1. No visible particle: no point
2. Particle difficult to be seen within 5 seconds: one point
3. Particle easily to be seen within 5 seconds: two points
4. Particles to be seen immediatly and in higher numbers: ten points

The formula for evaluation is: $Q_{TR} = \frac{A}{N}$, where A stands for 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 avalaible data is essential for the routine evaluation of the drug quality.

2.3 Meta-Analytic Modelling 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 = Number of Times with Q_{TR} positive in a series of n 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 1.

As has been pointed out in the literature (Pettiti [1994], Cooper and Hedges [1994]), the area of meta-analyis has received various impulses during its historic development. In psychology the development of measures were 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.*

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

Table 1. Number of defective items for 20 batches of 200 items each

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 paper, 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 for the data provided in Table 1 20 binomial tests are employed it can be expected that 1 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 consequence, investigators in quality assurance are forced to investigate for a non-existing source of deviation of the production process.

3 The Problem of Heterogeneity

3.1 The Occurrence of Heterogeneity

In fact, we are interested in separating *random deviations* which are occuring 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

[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 measurements and processes are done with the highest accurracy.

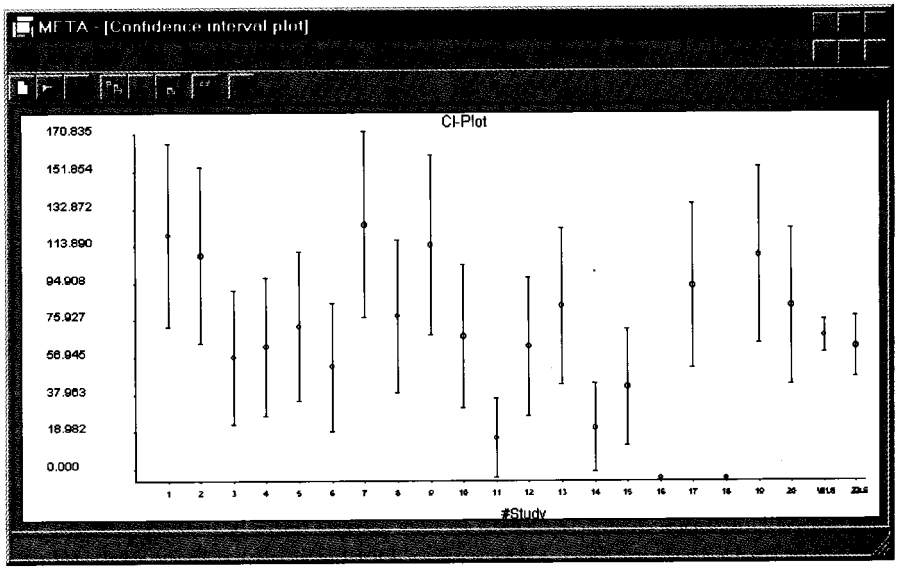


Fig. 1. Confidence interval plot from the package META for a textbook example of proprotion of defective items for 20 batches with 200 items each

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 $f(x,\theta) = \binom{m}{x}\theta^x(1-\theta)^{m-x}$, where m 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.

3.2 Diagnosis of Heterogeneity

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

$$\chi^2 = \sum_{i=1}^{n}(X_i - E(X_i))^2/Var(X_i).$$

Typically, $E(X_i)$ and $Var(X_i)$ 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 $E(X_i) = m\theta_i$ and $Var(X_i) = m\theta_i(1-\theta_i)$ which might lead

to the plug-in estimates $\widehat{E(X_i)} = X_i$ and $\widehat{Var(X_i)} = X_i(1-X_i/m)$. It can be shown that the associated distribution under homogeneity is quite different from a χ^2-distribution with $n-1$ df if sample sizes per batch, m, are small or moderate, even if the number of batches m becomes large. The right thing to do here turns out to be a variance estimate utilizing information from all batches: $\widehat{Var(X_i)} = S_n(1-S_n/m)$, where $S_n = \sum_{i=1}^{n} X_i/n$. The associated χ^2-statistic (with $\widehat{E(X_i)} = S_n$) can be shown to be validly approximated by a χ^2-distribution with $n-1$ df *even for small batch size* m (like $m = 5$). For further discussion see Böhning ([2000]) and Hartung ([1999]). To finish this aspect we find a value of $\chi^2 = 70.41$ with 19 df for the data of Table 1 which indicates strongly the presence of heterogeneity.

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

4 Modelling 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 $\theta_1 = \theta_2 = ... = \theta_n = \theta$ 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 $j = 1, ..., k$. It can be shown (Böhning [1999]) that then X_i has a mixture distribution

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

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

$$f(x_i, P) = \sum_{j=1}^{k} \binom{m}{x_i} \theta_j^{x_i} (1-\theta_j)^{m-x_i} p_j \tag{2}$$

To mention a second example, let us consider the effect measure of the Fisher-transformed correlation coefficient: $\theta = 0.5log[\frac{1+\rho}{1-\rho}]$. If we assume a normal distribution as conditional distribution for $\hat{\theta}_i$ with variance $1/(m_i-3)$ in the i-study, then the mixture distribution takes on the form

$$f(\hat{\theta}_i, P) = \sqrt{m_i - 3} \sum_{j=1}^{k} \varphi\{\sqrt{m_i - 3}(\hat{\theta}_i - \theta_j)\} p_j \tag{3}$$

where φ is the standard normal density.

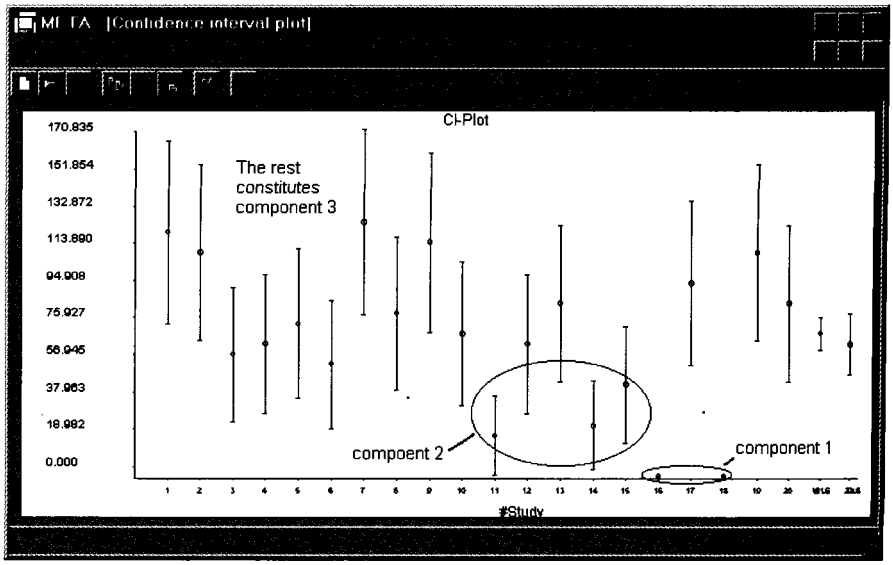


Fig. 2. Classification of the batches into their associated components for the textbook example of proprotion of defective items for 20 batches with 200 Items each

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 (2), 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 et al. ([1992],[1998]). The associated maximum likelihood estimate (NPMLE) of k and θ_j, p_j for $j = 1, ..., k$ 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 $k-1$, $k-2$, ... until no significant drop in the log-likelihood is notable. For these fixed values of k estimation is done via the EM-algorithm (Dempster et al. [1977]).

To provide a demonstration for this technique we again study the data of Table 1 and use the mixture model provided in (2). Tables 2 and 3 provide the results. There is empirical evidence for *heterogeneity* (see Table 2) and that this heterogenity consists of 3 components (Table 3).

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 and this components represents the majority of all batches (76.8%).

Number of components k	Log-likelihood
4 (NPMLE)	−63.1454
3	−64.0984
2	−70.9835

Table 2. Heterogeneity structure for 20 batches of 200 items each

Proportion θ_j	Weight p_j
0.0000	0.0996
0.0287	0.1326
0.0865	0.7678

Table 3. Estimated mixing distribution for $k = 3$

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

$$f(\theta|x_i) = \frac{f(x_i, \hat{\theta}_j)\hat{p}_j}{\sum_{l=1}^{k} f(x_i, \hat{\theta}_l)\hat{p}_l}, \text{ if } \theta = \hat{\theta}_j \text{ for some } j \in \{1, ..., k\} \qquad (4)$$

and $f(\theta|x_i) = 0$ otherwise, where $\hat{\theta}_j$ and $\hat{p}_j$ corresponds to the maximum likelihood estimate 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 $j = 1, ..., k$. This is done for the data in Table 1 and the results are provided in Table 4. Figure 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.

5 META – A Software Package for Meta-Analysis in Medicine, Social Sciences and the Pharmaceutical Industry

The software `META` has been developed to provide a tool which allows to perform meta-analyses within the areas of application described above. The focus of meta is on the analysis of heterogeneity, which may be considered here the unifying concept for several fields of application.

For different areas of application different measures of effect are important and necessary. Thus `META` enables the meta-analyst to choose out of a variety of measures of effect such as the relative risk in medicine, the standardized

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

Table 4. Classification of each batch into the components

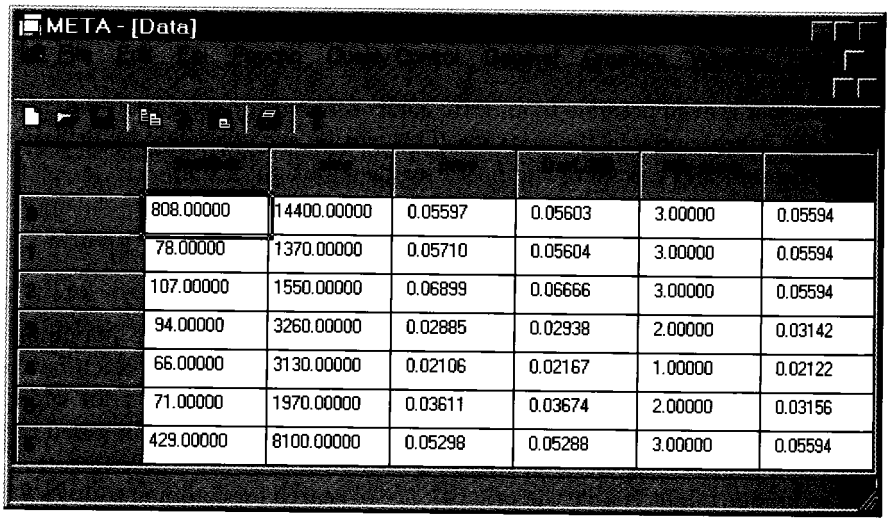


Fig. 3. Spreadsheet with original data and empirical Bayes estimates

difference in psychology and proportions in quality control, just to mention a few.

META provides various statistical methods to perform meta-analyses such as simple pooled estimates, random effects models and graphical procedures such has confidence interval plots, funnel plots etc. We will illustrate the possible use of META using a data set from psychiatric epidemiology.

5.1 A Worked Example from Psychiatric Epidemiology

The following meta-analysis investigates the prevalence of agoraphobia based on seven studies (Eaton, [1995]) in several countries all over the world. Ago-

raphobia may be defined as space anxiety, as a fear of being in public places. This psychiatric disorder may even lead to total avoidance of public places and thus may cause severe disability.

An initial step in any meta-analysis might be to plot the effect measure together with a 95-percent confidence interval. This may be done using **META** and its graphics facilities. Figure 2 shows a screen dump of **META** and its data window. The data window shows the prevalent cases of agoraphobia together with the population at risk of the respective study.

The simplest model possible assumes parametric density $f(x, \theta, \sigma^2)$ for some random quantity X where θ is a parameter of interest and σ^2 is a nuisance parameter which might or might not be present in the model. In the example at hand, $f(x, \theta) = \binom{m}{x}\theta^x(1-\theta)^{m-x}$. In this case all studies are assumed to measure the same overall effect θ and that they only differ in variability. Thus the summary measure needs to assign weights according to the inverse of the variance of the individual study in order to obtain the summary measure.

Looking at the confidence interval plot (Figure 5) there seems to be a large degree of variability to be present. However frequently one is interested in obtaining a summary measure for all studies. Using **META** we obtain the following results:

```
POOLED ESTIMATOR FOR PROPORTIONS

RESULTS
Pooled estimate:      0.048892
Common variance:    0.00000145
95 percent confidence interval (0.04654, 0.05125)

Chi-Square test for homogeneity of proportions:
     115.23539 df = 6  p-value:       0.00000
```

Clearly looking at the value of the χ^2-test of homogeneity we reject the null-hypothesis and conclude that there is substantial heterogeneity in terms of the prevalence of agoraphobia in the countries studied. As a result the computation of an overall rate is not very meaningful, since we ignore the underlying heterogeneity.

In order to deal with heterogeneity a mixture model has been implemented in **META**, as described in the previous sections. As it is assumed that θ is not constant, but is varying itself according to some further distribution P, we are able to consider the moments $E_P(\theta) = \mu$ and $Var_P(\theta) = \tau^2$ of the heterogeneity distribtuion P. Frequently τ^2 is called the *heterogeneity variance*. **META** offers modelling according to two different concepts in order to deal with heterogeneity: One is a moment approach and is based on equating the expected value of the χ^2-statistic to the oberserved one and the solve for τ^2. Actually this is the approach by DerSimonian and Laird ([1986]). The

DerSimonian-Laird-Estimator is provided by

$$\hat{\tau^2} = \frac{\chi^2 - (n-1)}{\sum_i w_i - \frac{\sum_i w_i^2}{\sum_i w_i}} \tag{5}$$

where $w_i = Var(\hat{\theta}_i)^{-1}$, the inverse of the variance of the measure of interest in the i-th study and $\chi^2 = \sum_i w_i(\hat{\theta}_i - \hat{\mu})^2$ with $\hat{\mu} = \sum_i w_i\hat{\theta}_i / \sum_i w_i$. $\hat{\tau^2}$ will only be computed if χ^2 is larger than $n-1$; otherwise it is set to zero. Having estimated τ^2 a pooled estimator is computed using the weights $w_i^* = \{Var(\hat{\theta}_i) + \hat{\tau^2}\}^{-1}$.

The other approach does find the nonparametric maximum likelihood estimator of the mixture model as outlined in section 4.

We proceed in our analysis with the estimation of the DerSimonian-Laird estimator:

```
RESULTS
Pooled estimate                                :        0.0455
(adjusted for heterogeneity using Dersimonian-Laird)

Heterogeneity variance:        0.0003
Variance of pooled estimator: 0.0000465

0.04545 95 percent CI: (0.0321,    0.0588)
```

Please note that we find a substantial value for the heterogeneity variance τ^2 in this data set. As expected incorporating heterogeneity leads to a larger variance for the DerSimonian-Laird estimator. As a result we obtain a much wider confidence interval compared to the pooled estimator where we assume a constant value for θ (see also Figs. 3 and 6).

Frequently there is a debate, whether one should use a summary measure in the presence of heterogeneity. One might argue that this may be done, but one has to be careful how to interpret the results. Under the presence of heterogeneity a summary measure will reflect the overall mean in the population well knowing that this effect might be different in subparts of the population.

If the presence of heterogeneity has been identified one might wish to model the *structure* of this heterogeneity and for example find the levels of effect in subparts of the population. This can be accomplished using the finite mixture model approach outlined above.

A convenient computational strategy uses a fixed grid of potential support points (subpopulation means θ_j) which may or may not receive weights p_j.

Figure 4 shows the dialog box which allows the user to define a grid of potential support points. Depending on the current measure of effect an appropriate mixing kernel may be chosen by the user. In this case since we are dealing with rates and the binomial distribution is the natural choice.

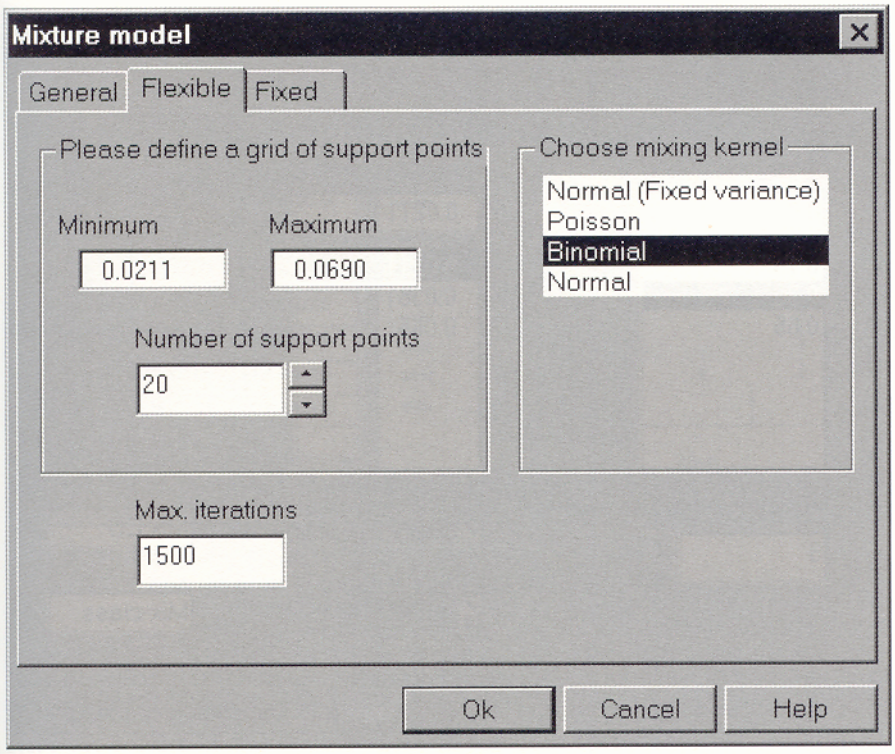


Fig. 4. Dialog box for flexible number of components mixture model

```
Initial number of components: 5
Parameter:    0.0211, Weight:    0.1441
Parameter:    0.0317, Weight:    0.2840
Parameter:    0.0530, Weight:    0.3073
Parameter:    0.0584, Weight:    0.1533
Parameter:    0.0690, Weight:    0.1113

Log-likelihood at iterate: -34.8009
```

Based on this grid `META` identifies five potential subpopulations. Now these grid points with positive support may be used to find a refined solution using the EM algorithm (Dempster, Laird and Rubin, [1977]). Here we keep the number of components fixed and update mixing weights and subpopulation means. Frequently some population means coincide and thus the number of components decreases. For our data at hand we find after applying the EM-algorithm four remaining components. (Results not shown here). Now a backward elimination approach may be used in order to reduce the number

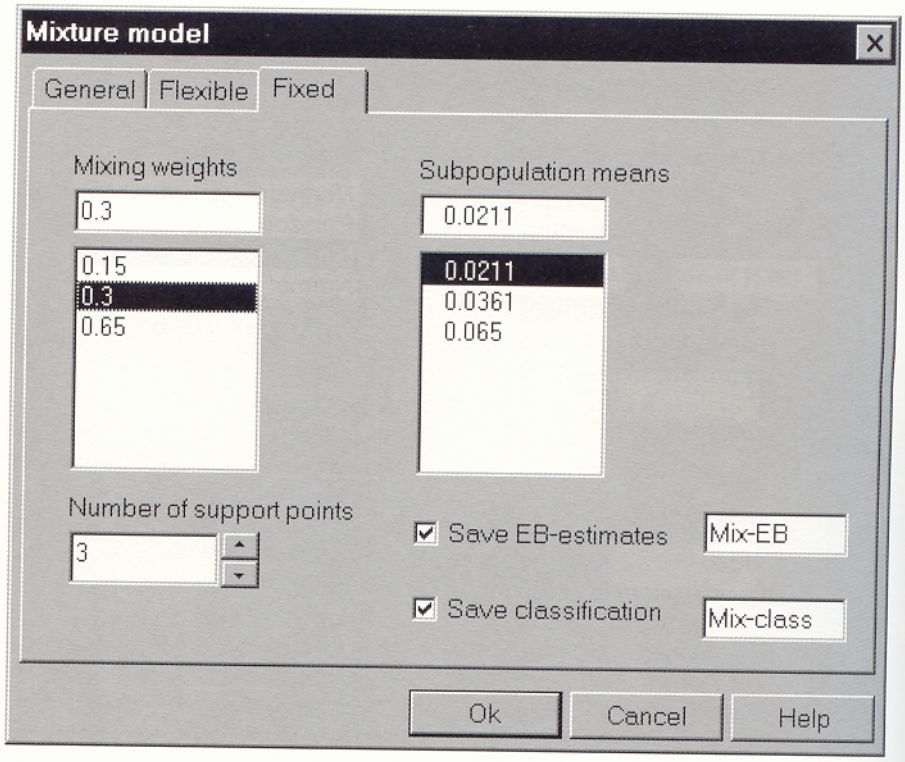


Fig. 5. Dialog box for fixed number of components mixture model

of mixing components. This would imply that we test $k = 4$ vs $k = 3$ using a Likelihood Ratio test approach.

```
NPMLE for Fixed support size
Number of components after combining equal parameter estimates: 3

Parameter:    0.0212, Weight:    0.1440
Parameter:    0.0316, Weight:    0.2844
Parameter:    0.0559, Weight:    0.5716

Log-likelihood at iterate: -34.3889
```

Clearly the log-likelihood is only slightly smaller for this three component mixture model and we would conclude that a three component solution is appropriate.

Once a mixture model has been chosen, one might be interested in classifying the individual study. Due to their discrete structure mixture models provide a natural way of classifying the individual study. This is achieved by applying Bayes theorem (see (4)). According to this rule the i-th study is

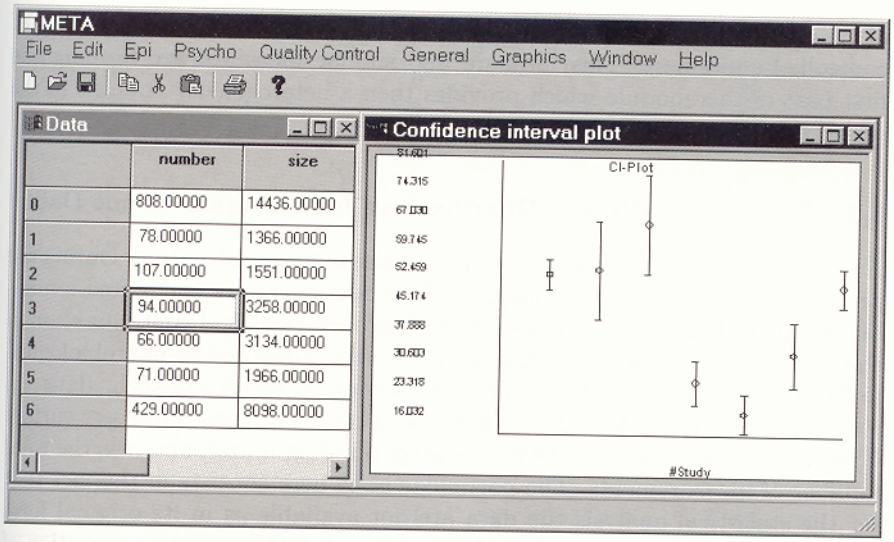


Fig. 6. Data window and confidence interval plot

then assigned to that subpopulation j for which it has the highest posterior probability of belonging. META offers the option to classify the studies and to store the results of this classification in the data spreadsheet.

META also computes the posterior expectation for the measure of effect for the individual study based on the assumed distribution. Likewise also the posterior expectations may be stored within the data frame.

5.2 Availability

META is designed to be platform independent and uses the wxWindows 2.0 class library (Julian Smart [2000]). META may be obtained for Microsoft Windows 9x/NT and for Unix(Linux) operating systems. META is available from the authors on request.

6 META's Special Module: *Quality Control*

6.1 Quality Measures for Parenteral Drugs

As has been outlined in Sect. 2.2, parenteral drugs are products which have to comply with additional, specific properties like sterility and essentially free of visible particles because of their parenteral application. Each parenteral container is controlled by a 100%-inspection for particulate matter. The particulate matter is evaluated for particles which can be seen easily, good or difficult, leading to the index $Q_{TR} = \frac{A}{N}$, where A stands for number

of points recorded by three test persons and N stands for the number of controlled containers. This measure and similar measures are forming the first part of this module which provides then a heterogeneity analysis in the sense of DerSimonian-Laird.

6.2 Estimating Mixture Distributions for Grouped Count Data

A more complete modelling is provided for the following situation. During the production process of a pharmaceutical product there are conducted a number of controlling measurements concerning bacteriological contamination in the air, on surfaces and on working clothes. This is denoted by microbiological inprocess controlling (MIPC). From this process result counting data w_i, $i = 1, \cdots, n$. Here w_i denotes the number of colony forming units per surface – respectively volume unit (CBU = colony building units), and n is the total number of measurements. The particular issue here consists in the fact that for the statistical analysis the data are not available as in its original form as exact count W_i, but as grouped data: there is a disjoint decomposition of the positive half axis given in the form

$$0 = r_0 < r_1 < \cdots < r_{L-1} < r_L := \infty\,,\ r_l \in \mathbb{N},\ 1 \le l \le L-1\,.$$

Furthermore we have the following notation: $R_l = \{r_{l-1}, \cdots, r_l - 1\}$, $l = 1, \cdots, L-1$, $R_L = \{r_{L-1}, r_{L-1} + 1, \cdots\}$. The observations available for the analysis are

$$n_l := \#\{i \in \{1, \cdots, n\} : w_i \in R_l\} \tag{6}$$

that means, n_l is the number of measurements for which the number of detected CBU falls within the group R_l. We have a loss of information: instead of the vector $\mathbf{w}= (w_1, \cdots, w_n)$ of original data the vector of 'observed' data is now $\mathbf{y} = (n_1, \cdots, n_L)$.

Let us consider an example from microbiological environmental monitoring. The raw data (Table 5) consist of the number of detected CBU per surface unit on the overalls of personal staff in an inprocess control department. Here, the total number of measurements is $n = 164$ with $n_0 = 84$ steril measurements(e.g. those samples with $w_i = 0$), and we have $L = 16$ groups.

The following heterogeneity model is implemented in our analysis. The variables W_i are independent distributed according to a mixture of Poisson distributions with mixing distribution P. This mixture distribution P is mixing k Poisson components $Po(\theta_j)$:

$$P(W_i = w|\theta_j) = \frac{\theta_j^w}{w!}\mathrm{e}^{-\theta_j}\,,\ w = 0, 1, \cdots \text{ with } \theta_j > 0\,,\ j = 1, \cdots k\,.$$

If we interprete the objects resp. individuals on which the measurements are done as representatives of a *heterogenuous* population and for which the

Group number l	Left bound for the group r_{l-1}	Number of CBU n_l
1	0	84
2	1	27
3	2	14
4	3	10
5	4	5
6	5	4
7	6	6
8	7	1
9	8	4
10	9	1
11	10	1
12	11	1
13	16	1
14	21	2
15	31	2
16	51	1

Table 5. CBU counts for personal overall in inprocess control department

subpopulation membership in unknown, then the mixture model arises as the marginal model where the margin is taken over the latent variable denoting the subpopulation membership. The population weight of the component number j is denoted by p_j with

$$0 < p_j \leq 1\,,\; j = 1, \cdots k\,,\; \sum_{j=1}^{k} p_j \;=\; 1\,.$$

The following quantities have to be determined resp. to be estimated:

1. k, the number of components,
2. the mixing fractions p_j,
3. the corresponding parameter θ_j, the means within the single components.

The mixture distribution model is of the following form:

$$f(w; k, \Psi) := P_{k,\Psi}\left(\{W = w\}\right) = \sum_{j=1}^{k} p_j f(w; \theta_j) = \sum_{j=1}^{k} p_j \frac{\theta_j^w}{w!} \mathrm{e}^{-\theta_j}\,, \qquad (7)$$

with $\Psi = (\theta_1, \cdots, \theta_k; p_1, \cdots p_{k-1})^T$. Combining (6) and (7) the estimation problem presented here exhibits two aspects:

1. There are to be estimated the parameter of a *mixture distribution*, for which
2. the data are at hand are of *grouped form*.

We use the maximum-likelihood principle for estimating the parameter Ψ. For this we have to begin with forming the likelihood function of the data $\mathbf{y}$, which are available. It appears, that the problem can be divided into two subproblems:
(i) The determination of the number k of mixture components and supplying initial estimations resp. starting values for part (ii).
(ii) The estimation of the parameter vector Ψ.
With regard to an algorithm for the numerical solution of the estimation problem it turns out, that the likelihood based on the 'observed' data $\mathbf{y}$, $l(\mathbf{y})$, is not much helpful. For example, within the score equation $(\partial/\partial p_j)l(\mathbf{y})$ it is not possible to separate the parameter p_j. Therefore we preferred a solution by means of the EM-algorithm via constructing a 'complete-data-likelihood' (see for instance McLachlan and Krishnan [1997]). META is using the derived iterative equations to sequentially approximate the maximum-likelihood estimation according to the EM-algorithm. These iterative equations are of the structure

$$\theta_j^{(t+1)} = \theta_j^{(t)} F_\theta(j, \Psi^{(t)}) \ , \ p_j^{(t+1)} = p_j^{(t)} F_p(j, \Psi^{(t)}) \, .$$

At this F_θ and F_p are expressions depending on the parameter values of the preceding step.
To solve the subproblem (i) we developed an initial algorithm. Within this algorithm initially we set $k = L$ and then generate a sequence of estimations $k_{(s)}, \Psi_{(s)} = (\theta_1, \cdots, \theta_{k(s)}; p_1, \cdots, p_{k(s)-1})^T$ which is stopped if a stoppping rule is met, say at step number $s = S_0$. Then, we set $k := k_{(S_0)}$ and $\Psi_{(S_0)}$ is the vector of starting values for the algorithm of subproblem (ii).

For our example of counted CBU on personal overalls the output of the algorithm is a mixing distribution with $k = 7$ components (see Table 6). Note that the weight is concentated on components with small mean value, nevertheless there are three components with very small weight but rather large mean value:

Component	Weight	Mean value
1	0.26176	0.00000
2	0.30838	0.40775
3	0.25031	1.71035
4	0.14184	5.92506
5	0.02016	22.4260
6	0.01269	43.4509
7	0.00486	59.5728

Table 6. Estimated mixing distribution for the example data in Table 5

In the light of the application we interpret the result as follows. It appears that the practitioner will focus attention on the components $5, 6, 7$ and, consequently, on the samples which are associated with these components, in order to identify potential sources of defiencies.

7 Discussion

We touched upon an approach which explicitly allows the modelling 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. Important measures of interest have been considered from the areas of medicine and pharmacy (relative risk and risk difference), psychology (standardized difference and correlation coefficient) and – as a new area of application – quality control and assurance (quality index and count index). For these three application areas different modules have been developed and assembled to form a package `META` which allows the user in a simple way to analyze data in his/her application. As a special feature of the package `META` heterogeneity analysis is provided for each application area on the basis of mixture modelling.

Important aspects of future research will be:

1. *valid* computation of standard errors of the parameters involved in the mixture model and their associated confidence intervals
2. extension of the mixture models to allow covariate modelling
3. inclusion of these aspects in the package `META`

At the current stage not much is known about the correct computation of standard errors for the parameters of a mixture model. Basford, Greenway, McLachlan and Peel ([1997]) compare two methods of getting standard errors for the parameters of a mixture model of *normal* distributions. One method used in their computation is based on the conventional method of inversion of the information matrix. The other is based on the technique of the Bootstrap (Efron and Tibshirani [1993]). Their competitive analysis of both methods provides evidence that – though the information theoretic (and less computational expensive) approach is frequently close to the Bootstrap method – the Bootstrap seems to provide a useful alternative, especially for extreme cases of mixture models and small sample sizes. In addition, the paper by Basford, Greenway, McLachlan and Peel ([1997]) provides an example with larger sample size, in which the information based standard error is in considerable disagreement with the Bootstrap standard error. Therefore, it will be investigated whether it is possible to implement the Bootstrap appraoch for the mixture models used in our application areas to provide valid standard errors and, thus by, confidence intervals.

Another important area of further research will be the question in which way covariates can be inlcuded into the mixture modelling. Often additional information is provided to be used in the explanation of heterogeneity found in the data. For example, quality control sample data might come from different departments, from different shifts of quality control workers, or might have been collected at varying points in time. All these pieces of information might be collected and might form covariates which can be investigated for their potential in explaining heterogeneity. If the heterogeneity can be fully explained by the observed covariates, then the sources of variation in the quality control measure or index has been captured. It will be of importance to allow for heterogeneity *given the covariates*: we call this *residual heterogeneity*. If there is no residual heterogeneity, search for further sources of variation will be superfluous, at least in the sample data. Therefore, it will be important to model residual heterogeneity, which can be only validly accomplished if the univariate mixture model is extended for covariates.

It is targetted to extend `META` for these additional complexities in the near future.

Acknowledgements. This research is done under support of the *Bundesmimisterium für Bildung und Wissenschaft, Forschung und Technologie (BMBF).*

References

[1997] Basford, K.E., Greenway, D.R., McLachlan, G.J., and Peel, D.: Standard errors of fitted component means of normal mixtures. *Computational Statisitcs* **12** (1997), 1–17.

[2000] Böhning, D.: The flaw in χ^2-heterogeneity tests: a revisit of the Neyman–Scott problem? *Unpublished Report, Free University Berlin* 2000.

[1999] Böhning, D.: *Computer-Assisted Analysis of Mixtures and Applications.* `META`*-Analysis, Disease Mapping, and Others.* Chapman & Hall/CRC 1999.

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

[1998] Böhning, D., Schlattmann, P. and Dietz, E.: Recent developments in computer-assisted analysis of mixtures (`C.A.MAN`). *Biometrics* **54** (1998), 367–377.

[1994] Cooper, H. and Hedges, L.: *The Handbook of Research Synthesis.* Russel Sage Foundation 1994.

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

[1997] Derman, C. and Ross, S.M.: *Statistical Aspects of Quality Control.* Academic Press 1997.

[1986] DerSimonian, R. and Laird, N.: Meta-Analysis in Clinical Trials. *Controlled Clinical Trials* **7**(1986), 177–188.

[1995] Eaton, W.W.: Progress in the epidemiology of anxiety disorders. *Epidemiologic Reviews* **17** (1995), 32–38.

[1993] Efron, B. and Tibshirani, R.J.: *An Introduction to the Bootstrap.* Chapman & Hall 1997.

[1999] Hartung, J.: An alternative approach in META-Analysis. In: Holling, H., Schulze, R. und Böhning, D. (eds): *Workshop Meta-Analysis, Münster 1999.*

[1985] Hedges, L.V. and Olkin, I.: *Statistical Methods for Meta-Analysis.* Academic Press 1985.

[1995] Jones, D.R.: Meta-analysis: weighing the evidence. *Statistics in Medicine* **14**(1995), 137–149.

[2000] Malzahn, U., Böhning, D., and Holling, H.: Nonparametric estimation of heterogeneity variance for the standardised difference used in meta-analysis. *Biometrika* **87** (2000), 619–632

[1997] McLachlan, G. J., Krishnan, T.: The EM Algorithm and Extensions John Wiley & Sons, Inc. 1997.

[1994] Petitti, D.B.: *Meta-Analysis, Decision Analysis and Cost-Effectiveness Analysis. Methods for Quantitative Synthesis in Medicine.* Oxford University Press 1994.

[1999] Normand, S.-L.N.: Tutorial in Biostatistics: Meta-analysis: formulating, evaluating, combining, and reporting. *Statistics in Medicine* **18**(1999), 321–359.

[2000] Smart, J.: *wxwindows – a cross-platform GUI-solution.* `Http://www.wxwindows.org` 2000.