Published in final edited form as:

J Appl Stat. 2012 ; 39(2): 445–460. doi:10.1080/02664763.2011.596193.

Two-stage hierarchical modeling for analysis of subpopulations in conditional distributions

Inna Chervoneva^{a,1}, Tingting Zhan^b, Boris Iglewicz^b, Walter W. Hauck^c, and David E. Birk^d

^aDivision of Biostatistics, Department of Pharmacology and Experimental Therapeutics, Thomas Jefferson University, Philadelphia, Pennsylvania 19107, U.S.A

^bDepartment of Statistics, Temple University, Philadelphia, Pennsylvania 19122, U.S.A

^cSycamore Consulting LLC, New Hope, PA 18938, U.S.A

^dDepartment of Pathology and Cell Biology, University of South Florida, Tampa, Florida 33612, U.S.A

Abstract

In this work, we develop modeling and estimation approach for the analysis of cross-sectional clustered data with multimodal conditional distributions where the main interest is in analysis of subpopulations. It is proposed to model such data in a hierarchical model with conditional distributions viewed as finite mixtures of normal components. With a large number of observations in the lowest level clusters, a two-stage estimation approach is used. In the first stage, the normal mixture parameters in each lowest level cluster are estimated using robust methods. Robust alternatives to the maximum likelihood estimation are used to provide stable results even for data with conditional distributions such that their components may not quite meet normality assumptions. Then the lowest level cluster-specific means and standard deviations are modeled in a mixed effects model in the second stage. A small simulation study was conducted to compare performance of finite normal mixture population parameter estimates based on robust and maximum likelihood estimation in stage 1. The proposed modeling approach is illustrated through the analysis of mice tendon fibril diameters data. Analyses results address genotype differences between corresponding components in the mixtures and demonstrate advantages of robust estimation in stage 1.

Keywords

Robust finite normal mixture; Weighted likelihood estimator; Hierarchical models; Mixed effects models; Two-stage estimation

1 Introduction

Nano-measures of various biological structures often yield hierarchically clustered data with large number of observations and multimodal distributions at the lowest level of clustering. One such example is considered here, with collagen fibril diameter measures collected from multiple animals with multiple microscopic fields per animal, and large samples of fibril diameters per microscopic field. From the biological point of view, it is of interest to study subpopulations of conditional distributions of fibril diameters and their dependence on fixed and random effects. Analyses of these data have been particularly useful in studies addressing regulation of hierarchical steps in collagen fibrillogenesis. This process involves

¹Corresponding author. i_chervoneva@mail.jci.tju.edu.

the initial assembly of protofibrils followed by linear and lateral growth of protofibrils to produce the mature fibril. Each step is regulated in a tissue-specific manner to generate the diversity of structure and function observed [5]. The subpopulation of the smallest diameters (first component of the normal mixture) includes the initially assembled immature protofibrils or fibril intermediates. The second component corresponds to the subpopulation of maturing fibrils, involving linear and lateral growth of preformed protofibrils [4]. The third component (not yet present in the data from 4 day old animals used as example here) represents the subpopulation of mature fibrils [25].

Clustered and correlated data are analyzed using mixed effects models, but standard (linear, generalized linear, and nonlinear) and semi-parametric mixed effects models focus on analysis of mean structure as a function of covariates, and do not identify or address properties of subpopulations. As an alternative, we propose a hierarchical model with conditional distributions at the lowest level of clustering modeled as finite mixtures of normal components. The lowest level cluster-specific means and log-transformed standard deviations are modeled in a linear mixed effects model. Our model may be viewed as an extension of the generalized linear finite mixture model [19], where finite normal mixture parameters are allowed to depend on fixed effects covariates, and of the finite mixture regression model with random effects [22], where also the component means, but not the variances may depend on random effects. In our model, both components means and variances depend on fixed and random effects, such as animal or microscopic field. To our knowledge, such hierarchical models have not yet been considered in the literature.

The fibril diameter data that motivates the proposed model framework come from studies of collagen fibril assembly during development in mice. Here, we analyze the tendon fibril diameter data. From each animal, tendons are cut perpendicular to the tendon and fibril axis. Then multiple fields are photographed using a transmission electron microscope. Negatives (5–6 per animal) are selected randomly, and fibril diameters measured in microscopic fields of defined size using an image analysis system. Each microscopic field yields a rather large sample of about 200–500 fibril diameter measurements. The accompanying Figure 1 presents histograms and kernel density estimates (Gaussian kernel) for distributions of tendon fibril diameters in sample microscopic fields from 4 day old wild type and decorin deficient mice. For these data, it is of interest to model distribution in each microscopic field as a mixture of two normally distributed components and to study the effect of genotype (wild type or decorin deficient) on the location and spread of the corresponding components.

Conditional distributions of the fibril diameters are often contaminated with outliers either due to genetic alterations (e.g. abnormally large fused fibrils) or cross sections through the tapered ends of fibrils (smaller than diameter in the main cylindrical part). Since it is well known that finite mixtures fitted via maximum likelihood can be very sensitive to outliers [16], robust alternatives to the maximum likelihood estimation are preferable to provide more stable results with minimized influence of outliers contaminating conditional distributions. Robust approach to estimating conditional finite mixture distributions is also plausible in situations where mixture components may exhibit moderate departures from the normal distribution other than outliers, but it is desirable to view them as normally distributed from the application point of view.

For estimation of the proposed hierarchical model, we use a two-stage approach, which is suitable for data with sufficiently large number of continuous measures per cluster in the lowest level of clustering [10]. This allows flexibility in choosing estimation methods independently for the first and second stages and also developing the second stage mixed effects models for possibly only a subset or transformed stage 1 parameter estimates. In the first stage, we use a robust estimator for the finite normal mixture models. For the tendon

fibril diameter data from 4 day old animals, 2-component normal mixtures are used, but the proposed hierarchical model can accommodate any a priori chosen number of components. The same number of components for all microscopic fields is appropriate from the biological point of view, which is attributing a common underlying distribution of the fibril diameters associated with each developmental age. Between-animal and between-microscopic field variability is viewed as coming from the small variation of animal ages and sampling variability due to shifts of cross-sections within tendons, respectively.

In Section 2, we formalize the hierarchical model with conditional finite normal mixture distributions and describe two-stage approaches to its estimation. Robust estimation of the finite mixture models in stage 1 is considered in Section 3. Section 4 describes the results of the small simulation study conducted to compare performance of finite normal mixture population parameter estimates based on robust and maximum likelihood estimation in stage 1. In Section 5, we apply the proposed hierarchical model to the tendon fibril diameters data set and compare analyses results utilizing the robust and maximum likelihood estimates of the finite mixture parameters. Section 6 concludes with a discussion.

2 Hierarchical model with conditional finite normal mixture distributions

Consider a hierarchical mixed effects model written in two stages. The first stage model defines within-cluster distribution at the lowest level of clustering (microscopic field in our data example), conditional on random effects. The second stage model describes between-cluster variation of conditional distribution parameters through a linear mixed effects model incorporating higher levels of clustering (animal in our example).

Let y_{ikj} be j^{th} response from the lowest level cluster $k = k(i)$ nested within the highest level

cluster i , $i = 1, \dots, M$, $k(i) = 1, \dots, m_i$, $j = 1, \dots, n_{ik}$, so that a total of $\sum_{i=1}^M \sum_{k=1}^{m_i} n_{ik}$ responses are observed. If there are intermediate levels of clustering, they are assumed reflected in the design matrices of random effects. Stage 1 model for responses y_{ikj} in cluster k is written as

$$\text{Stage 1: } y_{ikj} \sim f(y|\pi_{ik}, \theta_{ik}) = \sum_{l=1}^h \pi_{ikl} \varphi(y|\theta_{ikl}), \quad j=1, \dots, n_{ik}, \quad (1)$$

where $f(y|\pi_{ik}, \theta_{ik})$ is a h -component mixture of normal densities with weighting proportions π_{ikl} , means μ_{ikl} , and variances σ_{ikl}^2 , that is

$$\varphi(y|\theta_{ikl}) = \exp \left\{ -\frac{1}{2} [\ln(2\pi\sigma_{ikl}^2) + (y - \mu_{ikl})^2 / \sigma_{ikl}^2] \right\},$$

so that $\theta_{ikl} = [\mu_{ikl}, \sigma_{ikl}^2]^T$, $l = 1, \dots, h$, $\theta_{ik} = [\theta_{ik1}, \dots, \theta_{ikh}]$, $\pi_{ik} = [\pi_{ik1}, \dots, \pi_{ikh}]$, and $\sum_{l=1}^h \pi_{ikl} = 1$.

For each cluster $k(i)$, let us denote by $\mathbf{a}_{ik}$ a subvector of cluster-specific parameters, possibly transformed, which are to be modeled in stage 2 as dependent on population parameters. Let us denote by $\mathbf{a}_i = [\mathbf{a}'_{i1}, \mathbf{a}'_{i2}, \dots, \mathbf{a}'_{im_i}]^T$ the combined vector of cluster-specific parameters for

the highest level cluster i . The second stage model is postulated as a linear mixed effects model for these cluster-specific parameters:

$$\text{Stage 2: } \mathbf{a}_i = \mathbf{A}_i \boldsymbol{\beta} + \mathbf{B}_i \mathbf{b}_i, \quad \mathbf{b}_i \sim i.i.d. \mathbf{N}(\mathbf{0}, \mathbf{D}), \quad (2)$$

where $\boldsymbol{\beta}$ is a vector of fixed population parameters, $\mathbf{b}_i$ is a vector of random effects associated with the highest level cluster i , $\mathbf{A}_i$ and $\mathbf{B}_i$ are design matrices for the fixed and random effects, respectively, and $\mathbf{D}$ is the covariance matrix of the random effects.

For the tendon fibril diameters data, we are specifically interested in means and log transformed standard deviations of the normal components,

$\mathbf{a}_{ik} = [\mu_{ik1}, \ln \sigma_{ik1}^2, \dots, \mu_{ikh}, \ln \sigma_{ikh}^2]^T$. In a more general hierarchical model, the mixing proportions $\pi_{ik1}, \dots, \pi_{ik(h-1)}$, may also be included in the second stage, if desired. However, since the normal distribution assumption may not be appropriate even for transformed mixing proportions, a linear mixed effects model may not be general enough for such a case.

With the simple two-stage (STS) approach [9][10][18], the second stage linear mixed effects (LME) model (2) is fitted to the pseudo-data of cluster-specific parameters $\mathbf{a}_{ik}^*$, estimated in stage 1, as if they were true parameters $\mathbf{a}_{ik}$. The corresponding approximate LME model for the vector of estimated cluster-specific parameters for the highest level cluster i ,

$\mathbf{a}_i^* = [\mathbf{a}_{i1}^*, \mathbf{a}_{i2}^*, \dots, \mathbf{a}_{im_i}^*]^T$, is written as

$$\mathbf{a}_i^* \approx \mathbf{A}_i \boldsymbol{\beta} + \mathbf{B}_i \mathbf{b}_i. \quad (3)$$

This approach is called the simple two-stage (STS) method in the context of nonlinear mixed effects models [10]. It is expected to perform well when all n_{ik} are sufficiently large, so that the estimation errors in $\mathbf{a}_i^*$ may be assumed negligible as compared to the variability in $\mathbf{b}_i$. Implicitly, the estimation errors in $\mathbf{a}_i^*$ become absorbed into the variance components corresponding to random effects $\mathbf{b}_i$.

The level of uncertainty in estimating $\mathbf{a}_{ik}$ and the accuracy of approximation in (3) depend on the convergence rates for the stage 1 estimators and on the size n_{ik} of cluster $k(i)$. For smaller cluster sizes n_{ik} , the estimation errors in $\mathbf{a}_i^*$ may become comparable in magnitude to the variability of random effects $\mathbf{b}_i$. In such cases, one would expect to gain precision using the global two-stage (GTS) approach, which explicitly incorporates the uncertainty of estimating $\mathbf{a}_{ik}$. This method was described by Steimer et al. [18] and extended by Davidian and Giltinan [9] for nonlinear mixed effects models. Chervoneva et al. [7] have extended the GTS modeling approach by considering a general class of hierarchical models and using any square-root-n-consistent and asymptotically normal estimators from stage 1 as pseudo-data in the second stage LME model.

Thus, the GTS approach is applicable when the stage 1 estimators are known to be consistent and asymptotically normal as $n_{ik} \rightarrow \infty$,

$$\sqrt{n_{ik}}([\mathbf{a}_{ik}^* | \mathbf{a}_{ik}] - \mathbf{a}_{ik}) \xrightarrow{L} \mathbf{N}(\mathbf{0}, \sum_{ik}), \text{ as } n_{ik} \rightarrow \infty,$$

where Σ_{ik} is some finite covariance matrix. Then for sufficiently large n_{ik} ,

$\mathbf{a}_{ik}^* | \mathbf{a}_{ik} \approx \mathbf{N}(\mathbf{a}_{ik}, \mathbf{C}_{ik})$, where $\mathbf{C}_{ik} = n_{ik}^{-1} \sum_{i \in k} \mathbf{C}_i$ and the corresponding approximate LME model for $\mathbf{a}_i^*$ is written as

$$\mathbf{a}_i^* \approx \mathbf{A}_i \boldsymbol{\beta} + \mathbf{B}_i \mathbf{b}_i + \mathbf{e}_i^*, \quad (4)$$

where $\mathbf{e}_i^* \sim \mathbf{N}(\mathbf{0}, \mathbf{C}_i)$ approximately, $\mathbf{C}_i = \text{diag}\{\mathbf{C}_{i1}, \dots, \mathbf{C}_{im_i}\}$, $\mathbf{C}_{ik} = O(n_{ik}^{-1})$ element-wise, and $\mathbf{b}_l$ and $\mathbf{e}_i^*$ are independent for any $l, i = 1, \dots, M$. Chervoneva et al [7] considered conditional (on estimated $\mathbf{C}_{ik}$ from the first stage) restricted maximum likelihood estimation (CREML) of model (4) and established consistency and asymptotic normality of CREML estimators of population parameters in $\boldsymbol{\beta}$ and $\mathbf{D}$ as $\min_{ik} n_{ik}/M \rightarrow \infty, M \rightarrow \infty$ and under the standard regularity conditions for LME model (4). Implementation of GTS-CREML approach does not require any custom programming for the stage 2 estimation if the standard SAS PROC MIXED is available. However, the estimates of all $\mathbf{C}_{ik}$ have to be computed as a part of stage 1 estimation, and performance of the GTS-CREML approach relies on the accuracy of approximation $\mathbf{a}_{ik}^* | \mathbf{a}_{ik} \approx \mathbf{N}(\mathbf{a}_{ik}, \mathbf{C}_{ik})$. For the data example considered in this work and in simulations with similar hierarchical finite mixtures (data not shown), we did not find an advantage of incorporating matrices $\mathbf{C}_{ik}$ into second stage model using the GTS approach, most likely because of the large numbers of observations available per each microscopic field (200–500). The SAS macros computing matrices $\mathbf{C}_{ik}$ for 2- or 3-component finite normal mixtures may be requested from the first author.

3 Robust estimators of finite normal mixtures

The first stage model (1) may be estimated by fitting an h -component normal mixture

$f(y_{ikj} | \boldsymbol{\psi}_{ik}) = \sum_{l=1}^h \pi_{ikl} \varphi(y_{ikj} | \theta_{ikl})$, $\boldsymbol{\psi}_{ik} = [\pi_{ik1}, \dots, \pi_{ikh}, \theta'_{ik1}, \dots, \theta'_{ikh}]^T$, in each lowest level using either the maximum likelihood or robust estimation procedure. The maximum likelihood estimation is described in detail, for example, in McLachlan and Peel [16].

Robust approaches to fitting finite mixtures were considered by De Veaux and Kreiger [11], Cutler and Cordero-Brana [8], Markatou [14], and Fujisawa and Eguchi [12], among others. The general class of robust density power divergence estimators was proposed by Basu et al [2] for robust estimation of parameters from a general parametric family of models. These estimators minimize power divergence between the data-based empirical distribution function and a parametric model density. The class is indexed by a single parameter $\alpha > 0$ that controls the trade-off between robustness and efficiency. The maximum likelihood is the limiting case as α tends to zero. Fujisawa and Eguchi [12] further developed minimum density power divergence estimators for finite normal mixtures. Alternative robust minimum divergence estimation methods were developed for continuous data by Beran [3], Basu and Lindsay [1], and Markatou et al [15]. This approach, utilizing the symmetric χ^2 divergence, was further developed for fitting finite normal mixtures by Markatou [14]. Density power divergence estimators are simpler to implement than minimum divergence estimation methods, which require nonparametric estimates (usually kernel smoother) of the model based and empirical density. However, in our numerical studies reported as a part of our other work [24], in some settings, the minimum divergence estimators yielded more robust finite normal mixture parameter estimates than density power divergence estimators.

Following Basu et al [2], for fixed $\alpha > 0$, the minimum density power divergence (MDPD) estimator for parameter vector $\boldsymbol{\psi}_{ik}$ of the density function $f_{ik}(y_{ikj} | \boldsymbol{\psi}_{ik})$ for cluster k of size n_{ik} may be obtained by minimizing

$$d_{\alpha}(\psi_{ik}) = \int f(y|\psi_{ik})^{1+\alpha} dy - (1+\alpha^{-1})n_{ik}^{-1} \sum_{j=1}^{n_{ik}} f(y_{ikj}|\psi_{ik})^{\alpha}. \quad (5)$$

As α approaches zero, the minimum density power divergence converges to the Kullback-Leibler divergence, and the corresponding minimum divergence estimator is also the maximum likelihood estimator. Using $\alpha = 1$ yields the minimum integrated square error estimator [17]. We do not consider $\alpha > 0.5$, since the efficiency of the minimum density power divergence estimator decreases as α increases [2]. Minimizing (5) is also equivalent to maximizing

$$l_{\alpha}(\psi_{ik}) = \frac{1}{n_{ik}\alpha} \sum_{j=1}^{n_{ik}} f(y_{ikj}|\psi_{ik})^{\alpha} - \frac{1}{1+\alpha} \int f(y|\psi_{ik})^{1+\alpha} dy. \quad (6)$$

Expression (6) may be viewed as a modified likelihood. Fujisawa and Eguchi [12] demonstrated that unlike the usual likelihood, the modified likelihood (6) is bounded for the normal mixture model under the condition $\min_l \pi_{ikl} \geq \phi(\alpha)/n_{ik}$, where $\phi(\alpha) = \alpha^{-1} (1 + \alpha)^{3/2}$. The estimating equations for the minimum power divergence estimator are

$$\frac{1}{n_{ik}} \sum_{l=1}^{n_{ik}} u(y_{ikl}|\psi_{ik}) f(y_{ikl}|\psi_{ik})^{\alpha} - \int u(y|\mathbf{a}_{ik}) f(y|\psi_{ik})^{1+\alpha} dy = 0,$$

where $u(y|\psi_{ik}) = \frac{\partial}{\partial \psi_{ik}} \ln f(y|\psi_{ik})$ is the maximum likelihood score vector for a h -component

normal mixture $f(y|\psi_{ik}) = \sum_{l=1}^h \pi_{ikl} \varphi(y|\theta_{ikl})$. It follows from the results of Basu et al. [2] that under the assumed model $f(y|\psi_{ik})$, the MDPD estimator of ψ_{ik} is an M-estimator, which is asymptotically multivariate normal with the covariance matrix $M^{-1}QM$, where

$$\begin{aligned} M &= \int u(y|\psi_{ik}) u(y|\psi_{ik})^T f(y|\psi_{ik})^{1+\alpha} dy \\ Q &= \int u(y|\psi_{ik}) u(y|\psi_{ik})^T f(y|\psi_{ik})^{1+2\alpha} dy - \xi \xi^T \\ \xi &= \int u(y|\psi_{ik}) f(y|\psi_{ik})^{1+\alpha} dy. \end{aligned} \quad (7)$$

To compute density power divergence estimates for each lowest level cluster $k = k(i)$, we implement the EM-like algorithm of Fujisawa and Eguchi [12]. The starting parameter values are computed as described in Woodward et al [21]. The approximate covariance matrices $\mathbf{C}_{ik}$, necessary for fitting the second stage GTS model (4), are computed substituting the final parameter estimates into (7).

The choice of the tuning parameter α may be adaptive to each sample, as in [12] or fixed a priori based on desired efficiency and robustness trade-off. For our application, we choose to select one common parameter α guided by the objective to minimize the between-cluster variability in stage 1 parameter estimates. This approach is consistent with the biological hypothesis of a common underlying distribution of the fibril diameters associated with each developmental age and genotype, and assumption that between-cluster variability is reflective of the experimental sampling variability.

Alternative previously proposed robust estimator of finite normal mixture is the weighted likelihood (WLEE) estimator [15][14]. The weighted likelihood estimating equations are

$$\sum_{j=1}^{n_{ik}} w(\delta(y_{ikj})) [\nabla_{\psi} \ln f(y_{ikj} | \psi_{ik})] = 0, \quad (8)$$

where $\delta(y_{ikj}) \in [-1, \infty]$ is the Pearson residual evaluated at y_{ikj} and $w(\cdot)$ is a weighting function. Pearson residual is defined by

$$\delta(y_{ikj}) = \frac{g^*(y_{ikj})}{f^*(y_{ikj} | \psi_{ik})} - 1,$$

where $g^*(y_{ikl})$ is the kernel density estimator of unknown density function cluster in $k(i)$ and $f^*(y_{ikj} | \psi_{ik})$ is the mixture model smoothed using the same kernel. For fitting finite normal mixtures, Markatou [14] proposed using $w(\delta) = 1 - (\delta/(\delta + 2))^2$, which corresponds to the symmetric chi-square distance. The solution of system (8) is called WLEE estimator of

$\psi_{ik} = [\pi_{ik1}, \dots, \pi_{ikh}, \theta'_{ik1}, \dots, \theta'_{ikh}]^T$. System (8) may also be rewritten as separate subsystems for parameters $\pi_{ik1}, \dots, \pi_{ikh}$ [14],

$$\pi_{ikl} = \sum_{j=1}^{n_{ik}} \frac{w(\delta(y_{ikj}))}{\sum_{j=1}^{n_{ik}} w(\delta(y_{ikj}))} \tau_{ikjl}, \quad \tau_{ikjl} = \frac{\pi_{ikl} \varphi(y_{ikj} | \theta_{ikl})}{f(y_{ikj}, \theta_{ikl})}$$

and for parameters θ'_{ikl} ,

$$\sum_{j=1}^{n_{ik}} \sum_{l=1}^h w(\delta(y_{ikl})) u(y_{ikj}, \theta_{ikl}) \tau_{ikjl} = 0,$$

where $\tau_{ikjl} = \pi_{ikl} \varphi(y_{ikj} | \theta_{ikl}) / f(y_{ikj}, \theta_{ikl})$ and $u(y_{ikj}, \theta_{ikl}) = \nabla_{\theta_{ikl}} \ln \varphi(y_{ikj} | \theta_{ikl})$.

Markatou [14] proposed generating starting values using bootstrap subsamples from the data to compute method of moments estimates. For normal mixtures, the normal density with the variance h^2 was proposed as a natural kernel with the bandwidth parameter

$h^2 = c \sum_{l=1}^h \widehat{\pi}_{ikl} \widehat{\sigma}_{ikl}^2$, where c is a solution of equation

$$-A/2 \left\{ \left[\frac{(1+c)^{3/2}}{c(c+3)^{1/2}} \right]^{\dim(\psi_{ik})} - 1 \right\} = \gamma_0,$$

where $\gamma_0 > 0$ is the number of observations to be downweighed on average, selected by the user. Since we did not have a priori knowledge of expected number of outliers to be downweighed on average, the bandwidth parameter was computed as $h = \text{MAD}(x) / \sqrt{n_{ik}}$, where MAD is the median absolute deviation. This approach was developed in [24], and it does not require hypothesizing the number of potential outliers.

4 Simulation study

A small simulation study was conducted to compare performance of the population parameter estimates based on robust and maximum likelihood estimation of the finite normal mixtures in stage 1. One thousand (1000) data sets were simulated with the hierarchical and distributional structure similar to the data example. That is, we simulated 6 animals in each of the 2 groups/genotypes ($M = 12$) with 5 microscopic fields per each animal ($m = 5$), and $n_{ik} = 150$ observations per microscopic field. In each microscopic field, conditional distribution of fibril diameters was either uncontaminated 2-component normal mixture distribution, or included either left side or right side contamination. For each microscopic field k from animal i , the uncontaminated data were generated as n_{ik} realizations of the random variable $X_{ik} = P_{ik}X_{ik1} + (1 - P_{ik})X_{ik2}$, where $X_{ikc} \sim N(\mu_{ikc}, \sigma_{ikc}^2)$, $c = 1, 2$, $P_{ik} \sim \text{Bernoulli}(0.45)$, and

$$\begin{aligned}\ln\sigma_{ik1} &= 1.0 + \tau_{ik1}, \quad \tau_{ik1} \sim N(0, 0.1^2) \\ \ln\sigma_{ik2} &= 1.0 + \tau_{ik2}, \quad \tau_{ik2} \sim N(0, 0.12^2) \\ \mu_{ik1} &= -7.5 + \xi_{ik1}, \quad \xi_{ik1} \sim N(0, 3.5^2) \\ \mu_{ik2} &= 7.5 + \xi_{ik2}, \quad \xi_{ik2} \sim N(0, 5.5^2)\end{aligned}$$

The vectors of microscopic field random effects $[\tau_{ik1}, \tau_{ik2}, \xi_{ik1}, \xi_{ik2}]^T$ were also assumed to have the following correlation matrix:

$$\begin{pmatrix} 1 & 0.9 & 0.8 & 0.5 \\ 0.9 & 1 & 0.7 & 0.5 \\ 0.8 & 0.7 & 1 & 0.6 \\ 0.5 & 0.5 & 0.6 & 1 \end{pmatrix}$$

Two contamination scenarios were used to compare robustness of alternative methods under contaminations similar to the real life situations. For the left side contamination, 5% of the data in each microscopic field, originally generated under the main model, were replaced with the data generated from the third normally distributed contamination component

distributed as $N(\mu_{ik1} - 10, \sigma_{ik1}^2)$. This contamination mimics the real situation when some of the fibrils are cut through the tapered ends and corresponding measured diameters may be substantially lower than diameters of the main cylindrical part. For the right side contamination, 5% of the data generated under the main model were replaced with measurements uniformly distributed on the interval $[\mu_{ik2} + 3\sigma_{ik2}, \mu_{ik2} + 3\sigma_{ik2} + 20]$. This contamination mimics abnormally large diameters, possibly resulted from the fibril fusion or uncontrolled growth in genetically altered mice strains.

In the first stage of hierarchical modeling, 2-component normal mixture model was fitted to each microscopic field using the maximum likelihood (ML), minimum density power divergence (MDPD), and symmetric chi-square WLEE estimators, respectively. Then, the second stage model was estimated as an LME model $a_{ik} = \beta + b_{ik}$, where $a_{ik} = [\ln \sigma_{ik1}, \ln \sigma_{ik2}, \mu_{ik1}, \mu_{ik2}]^T$, $\beta = [\ln \sigma_1, \ln \sigma_2, \mu_1, \mu_2]^T$ is a 4×1 vector of population average finite mixture parameters, and b_{ik} , $k = 1, \dots, m$, is a 4×1 vector of random effect of microscopic field, such that $b_{ik} \sim N_4(\mathbf{0}, \mathbf{D})$ and $\mathbf{D}$ is unstructured positive definite 4×4 matrix. This simulation study was conducted in R. All the first and the second stage models (using the `lme()` function) converged regardless of the estimation method. For comparison, we have also included an alternative “naive” or basic modeling approach for analysis of subpopulations in 2-component finite mixture. The basic approach fits a 2-component finite mixture model to

all data pooled from all microscopic fields and ignoring the clustered structure and random effects in the finite mixture parameters. This basic approach mimics the way this kind of data were presented in cell biology publications before the hierarchical model with finite normal mixtures was proposed. For fitting such 2-component finite mixture models to pooled data, we used the maximum likelihood (ML), minimum density power divergence (MDPD), and symmetric chi-square WLEE estimators.

Results of estimating population finite mixture parameters are presented in Figure 2. The box plots summarize the observed relative % bias in each of the finite mixture parameters ($\mu_1, \mu_2, \ln \sigma_1, \ln \sigma_2$). For each parameter, the first boxplot corresponds to ML estimates, the second one corresponds to MDPD estimates, and the third one corresponds to WLEE estimates. In the left panel of plots, ML, MDPD, and WLEE refers to the estimation method in stage 1. In the right panel of plots, ML, MDPD, and WLEE refers to the method of fitting one 2-component finite mixture model to the pooled data ignoring clustering. The results for uncontaminated data (top row in Figure 2) suggest that without outliers, the accuracy and precision of estimating population finite mixture parameters are similar for robust and ML estimation in stage 1 while an appropriate hierarchical model is employed. Meanwhile, the “naive” basic approach results in generally biased estimates for all parameters ($\mu_1, \mu_2, \ln \sigma_1, \ln \sigma_2$). As may be expected, the bias is much more severe for estimating the log standard deviations, since these estimates get inflated from absorbing variability of the random effects. Similar pattern of relative bias in “naive” basic approach estimates is observed also in both contaminated scenarios with relatively smaller effects of contamination.

In contrast, results from the hierarchical models exhibit increases in relative bias consistent with the introduced contamination. For the scenario with the left side contamination, the estimates of the first component parameters (μ_1 and $\ln \sigma_1$) are generally biased. However, the bias is largest while using the ML stage 1 estimates, and the smallest while using the WLEE stage 1 estimates. For the scenario with the right side contamination, the estimates of the second component parameters are biased when based on the ML and MDPD stage 1 estimates (more heavily for ML estimates, as expected) and essentially unbiased when based on the WLEE stage 1 estimates. Slightly different results for the left and right side contaminations based on WLEE stage 1 estimates suggest that WLEE more efficiently down weighs the more extreme outliers. Overall, our simulation study demonstrates that most accurate estimation of the population finite mixture parameters is achieved by using an appropriate hierarchical model (rather than “naive” basic approach ignoring clustering) with the WLEE stage 1 estimation.

5 Analysis of the tendon fibril diameter data

The tendon fibril data were collected to study an effect of a specific genetic mutation, decorin-deficiency, on the process of collagen fibrillogenesis. The tendon fibril diameters were measured in 6 wild type and 6 decorin-deficient 4 day old animals with 5 microscopic fields per animal. Each microscopic field yielded a sample of 204–528 fibril diameters with the median of 360 observations. Figure 1 shows histograms and kernel density estimates (Gaussian kernel) for conditional distributions in sample microscopic fields. Generally bimodal distributions support the biological hypothesis of two subpopulations of fibril diameters for four day mice. The first subpopulation is viewed as a subpopulation of **small diameter protofibrils**, which is present in different proportions in all ages [25]. The second is a subpopulation of actively growing and/or fusing together fibrils. Therefore, we model the tendon fibril diameters data using model (1) with 2 normal components.

The goal of our analysis is to compare the means and the standard deviations of the corresponding subpopulations (components in the mixture) between wild type and decorin-

deficient animals. The variability in mixing proportions is viewed as reflection of the small variability in animal ages and sampling variability from different parts of the tendon. It is biologically plausible to assume that subpopulations have normal distributions. Meanwhile, using robust estimators of conditional finite normal mixtures should provide less biased estimates of location and scale of the mixture components even if their distributions are not quite normal.

In stage 1, model (1) with 2 normal components was fitted using the maximum likelihood (ML), minimum density power divergence (MDPD), and symmetric chi-square WLEE estimators. For the MDPD estimator, the tuning parameter α was selected from 0.2, 0.25, 0.3 by minimizing the Cramervon Mises divergence, as described in Fujisawa and Eguchi [12]. Smaller tuning parameter $\alpha = 0.1$ was also considered, but the resulting MDPD estimators were very close to the corresponding ML ones. Larger α was not considered since substantial efficiency loss is expected [2]. For the symmetric chi-square WLEE estimators, the band-width parameter was $h = MAD(x) / \sqrt{n_{ik}}$ for the microscopic field with n_{ik} observations. The finite normal mixtures were fitted in R. Convergence was achieved in all 60 available microscopic fields for all stage 1 estimators.

Figure 3 shows microscopic field specific means and log transformed standard deviations of 2-component normal mixtures estimated in stage 1 using the MLE, MDPE and WLEE estimators. The random effects of animal appear small in component means and more pronounced in log transformed standard deviations. Noticeably, the MDPE estimates are much closer to the ML estimates than the WLEE estimates. Furthermore, for the lower first component means, which are more likely to be downward biased as a result of lower end outliers, the WLEE estimates are noticeably higher than MLE and MDPE estimates (Figure 3). This suggests that lower end outliers (expected as a result from cutting through the tapered ends of the fibrils) are better down weighted by the WLEE estimates.

In the second stage, microscopic field specific means and log transformed standard deviations of 2-component normal mixtures, $\mathbf{a}_{ik}^* = [\ln \hat{\sigma}_{1ik}, \ln \hat{\sigma}_{2ik}, \hat{\mu}_{1ik}, \hat{\mu}_{2ik}]^T$, were modeled as dependent on genotype $g = 1, 2$, body weight w_i of animal i , and also on random effects of animal i and microscopic field $k = 1, \dots, m$ ($m = 5$),

$$\mathbf{a}_{ik(g)}^* = \beta_g + \beta_w w_i + \mathbf{b}_{i0} + \mathbf{b}_{ik}, \quad (9)$$

where the vectors of population parameters were

$$\beta_g = [\ln \sigma_{1g}, \ln \sigma_{2g}, \mu_{1g}, \mu_{2g}]^T, g = 1, 2, \text{ and } \beta_w = [\beta_{w\sigma 1}, \beta_{w\sigma 2}, \beta_{w\mu 1}, \beta_{w\mu 2}]^T.$$

Figure 4 indicates positive correlation between the means of components 1 and 2, between log standard deviations of components 1 and 2, and between mean and log standard deviation for each component. Therefore initial model for the covariance structure included uncorrelated random effects of animal in each mixture parameter and unstructured covariance matrix for the random effects of microscopic field. That is, it was assumed that vector of random animal effect $\mathbf{b}_{i0} = [b_{i0}^{\sigma 1}, b_{i0}^{\sigma 2}, b_{i0}^{\mu 1}, b_{i0}^{\mu 2}]^T$ has normally distributed independent components, $\mathbf{b}_{i0} \sim \mathbf{N}_4(\mathbf{0}, \text{Diag}(\tau_{\sigma 1}^2, \tau_{\sigma 2}^2, \tau_{\mu 1}^2, \tau_{\mu 2}^2))$ and vector of random effect of microscopic field $\mathbf{b}_{ik} = [b_{ik}^{\sigma 1}, b_{ik}^{\sigma 2}, b_{ik}^{\mu 1}, b_{ik}^{\mu 2}]^T$ has 4-variate normal distribution $\mathbf{b}_{ik} \sim \mathbf{N}_4(\mathbf{0}, \Sigma)$, where Σ is some positive definite 4×4 matrix. It is not really feasible to estimate unstructured

covariance matrix for the random effects of animal for these data with only 12 animals. The second stage LME models were fitted in SAS 9.2 (SAS Institute Inc., Cary, NC, USA).

After fitting the initial stage 2 model (9) to ML, MDPD, and WLEE parameter estimates, the parsimonious model was selected using the BIC statistics and significance tests for the fixed effects parameters, as described in Verbeke and Molenberg [20]. Using either WLEE or MDPD robust stage 1 estimates, the weight of animal had significant effect on log standard deviations but not on the means of the normal components. Using the ML stage 1 estimates, the weight of animal had significant effect on the log standard deviation of the second component only, while for the first component, the weight effect was only borderline significant. To allow comparison of results based on alternative stage 1 estimates, the second stage model was selected to be the same with all three types of stage 1 estimates. Thus, the final parsimonious stage 2 model was constrained to $\beta_{w\mu 1} = \beta_{w\mu 2} = 0$ for all of stage 1 estimates. Estimated from initial stage 2 models covariance matrices Σ included all significant positive correlation coefficients ranging from 0.32 to 0.94. All second stage models indicated no random effect of animal present in the first component log standard deviation (LME model yielded zero estimate for $\tau_{\ln\sigma 1}^2$). Meanwhile, animal random effect variance components $\tau_{\ln\sigma 2}^2, \tau_{\mu 1}^2, \tau_{\mu 2}^2$, were not estimated zero, but substantially smaller in magnitude than corresponding variance components of the microscopic field random effects.

Even these animal random effect variance components ($\tau_{\ln\sigma 2}^2, \tau_{\mu 1}^2, \tau_{\mu 2}^2$) were not always significantly different from zero (using Wald test available through the covtest option in SAS PROC MIXED), they were retained in stage 2 models to control for animal random effects, and non-significance was considered potentially attributed to the small number of animals in the data set. Table 1 reports fixed effects parameter estimates obtained by fitting the final second stage STS model to ML, MDPD, and WLEE stage 1 estimates, as well as the p-values for the t-tests of the genotype differences in each finite mixture parameter. The degrees of freedom for t-tests and F-tests were computed using the approach of Kenward and Roger [13]. Table 2 presents the covariance parameter estimates from the final second stage models. Examination of residuals and BLUPs of animal random effects did not indicate violations of the normal distribution assumption with any stage 1 estimates.

Parameter estimates of average component means from the second stage models were very similar for all stage 1 estimates, but the standard errors were generally lower using the robust (MDPD or WLEE) than ML estimation of stage 1 model (Table 1). The test results based on MDPD or WLEE estimates indicate that means of both normal components are significantly larger for the mutant than for wild type animals, while results based on ML estimates identify only the means of the first normal components as significantly different between the mutant and wild type animals. The estimates of the average log standard deviations of normal components were slightly lower for the robust than for ML stage 1 estimates (Table 1), which is consistent with down weighting potential outliers by the robust methods. Covariance parameter estimates (elements of matrix Σ) were generally similar based on all stage 1 estimates, but their standard errors were mostly higher for ML estimation approach and similar for two robust approaches in stage 1. This is consistent with more between-microscopic fields variability of stage 1 ML than robust estimates (Figure 1). The GTS approach was also implemented for ML and MDPD stage 1 estimates and yielded virtually the same results as the STS approach for the same stage 1 estimation approach (results are not shown).

In conclusion, the proposed hierarchical model with conditional 2-component normal mixture distributions provided necessary statistical framework to evaluate the effect of a certain genetic mutation on subpopulation of fibrils in the tendons of mice. Analysis results indicate a significant increase in diameters of preformed intermediates (first component in

the mixture) and in diameters of actively growing fibrils (second component in the mixture) due to decorin deficiency. Meanwhile, the variability of fibril diameters in either mixture component was not affected by this mutation. This is not necessarily the case with other mutations (e.g. [6]). The use of robust MDPD or WLEE instead of ML estimates in stage 1 improved precision of estimating population parameters in stage 2. This allowed identification of the biologically important difference between the second component means as significant in addition to the difference between the first component means also identified by using stage 1 ML estimates. This supports biological hypothesis that decorin deficiency further manifests in the process of fibril growth.

6 Discussion

In this work, we consider a modeling approach for multi-level clustered data where conditional distributions at the lowest level of clustering are viewed as finite mixtures of a priori known number of normal components and the interest is in analysis of subpopulations/components. Such data are common in studies that involve nano-measures of various biological structures, when moderate number of animals is used, but a large number of observations per animal is collected in clusters that correspond to microscopic fields. The proposed framework allows flexible modeling of means and standard deviations of subpopulations as functions of fixed and random effects. Models with conditional finite normal mixtures have been considered before [19][22], but not with both means and variances of normal components depending on fixed and random effects.

The adapted two-stage estimation approach provides high flexibility in selecting estimation methods in each stage independently. In this work, maximum likelihood was used in the second stage, while robust alternatives were utilized in the first stage. Robust estimation in stage 1 alleviates the problem with high sensitivity to potential outliers, which is known for maximum likelihood estimation of the finite normal mixtures. In addition, such robust estimation makes the model useful when distributions of mixture components may exhibit moderate departures from the normal distribution other than outliers. However, the two-stage approach is suitable only for the data with sufficiently large number of observations at the lowest level of clustering.

We considered the number of components in each conditional distribution mixture to be known a priori from the context of the applied problem. When this is not the case, the problem of determining the number of components is known to be difficult. Using robust approach to modeling a finite mixture adds additional challenge since the same observations may be viewed as either outliers or as a legitimate component in the mixture. To our best knowledge, the choice of the number of components for robust modeling of finite mixtures has not yet been addressed even for a one simple random sample problem. In the context of hierarchical data like ours, the unknown number of components may potentially vary between conditional distributions. This adds yet another level of complexity, which is beyond the scope of this paper. Therefore, the problem of determining the number of components will be addressed in further research.

In this work, we used robust methods only in stage 1, where outliers and some violations of model assumptions were expected. Meanwhile, normal distribution assumptions were appropriate for the stage 2 model. When necessary, it is straightforward to extend the two-stage approach to incorporate robust estimation of the stage 2 model, similar in spirit to the work of [23]. Furthermore, the two-stage approach allows extensions incorporating a non-linear or semiparametric mixed effects model as a stage 2 model.

The proposed hierarchical model with conditional finite normal mixture distributions was fitted to the tendon fibril diameters data to evaluate the impact of a specific mutation

(decorin deficiency) on the means and variances of the normal components. This analysis framework helps to advance our understanding of the mechanisms regulating collagen fibrillogenesis in specific tissues. In particular, significant increase in diameters of both preformed intermediates and actively growing fibrils due to decorin deficiency indicates that decorin is a key regulatory molecule in fibrillogenesis and tendon development. For comparison, the tendon fibril diameters from wild type and decorin deficient mice were also analyzed using the maximum likelihood estimation of finite mixture parameters. Analyses results demonstrates advantages of the robust stage 1 estimation, which increases precision and provides superior ways for detecting genotype differences in individual mixture components.

Acknowledgments

This work was supported by the grants AR054596 and AR44745 from NIH/NIAMSD. Tingting Zhan was supported by a Merck Quantitative Research Fellowship to Temple University. Boris Iglewicz was supported while on a Temple University Study Leave.

References

1. Basu A, Lindsay BG. Minimum disparity estimation for continuous models: efficiency, distributions and robustness. *Annals of the Institute of Statistical Mathematics*. 1994; 48:683–705.
2. Basu A, Harris IR, Hjort NL, Jones MC. Robust and efficient estimation by minimising a density power divergence. *Biometrika*. 1998; 85:549–559.
3. Beran R. Minimum Hellinger distance estimates for parametric models. *The Annals of Statistics*. 1977; 5:445–463.
4. Birk DE, Nurminskaya MV, Zycband EI. Collagen fibrillogenesis in situ: fibril segments undergo post-depositional modifications resulting in linear and lateral growth during matrix development. *Developmental Dynamics*. 1995; 202:229–243. [PubMed: 7780173]
5. Birk, DE.; Bruckner, P. Collagens, suprastructures and collagen fibril assembly. In: Mecham, RP., editor. *The Extracellular Matrix: an Overview*. Vol. 1. Springer; NY: 2011. p. 77-115.
6. Ezura Y, Chakravarti S, Oldberg Å, Chervoneva I, Birk DE. Differential expression of lumican and fibromodulin regulate collagen fibrillogenesis in developing mouse tendons. *Journal of Cell Biology*. 2000; 151:779–788. [PubMed: 11076963]
7. Chervoneva I, Iglewicz B, Hyslop TA. General Approach for Two-Stage Analysis of Multi-Level Clustered Non-Gaussian Data. *Biometrics*. 2006; 62:752–759. [PubMed: 16984317]
8. Cutler A, Cordero-Brana OI. Minimum Hellinger distance estimation for finite mixture models. *Journal of the American Statistical Association*. 1996; 91:1716–1723.
9. Davidian M, Giltinan DM. Some simple methods for estimating intraindividual variability in nonlinear mixed effects models. *Biometrics*. 1993; 49:59–73.
10. Davidian, M.; Giltinan, DM. *Nonlinear Models for Repeated Measurement Data*. Chapman and Hall; London: 1995.
11. De Veaux RD, Krieger AM. Robust estimation of a normal mixture. *Statistics & Probability Letters*. 1990; 10:1–7.
12. Fujisawa H, Eguchi S. Robust estimation in the normal mixture model. *Journal of Statistical Planning and Inference*. 2006; 136:3989–4011.
13. Kenward MG, Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*. 1997; 53:983–997. [PubMed: 9333350]
14. Markatou M. Mixture models, robustness, and the weighted likelihood methodology. *Biometrics*. 2000; 56:483–486. [PubMed: 10877307]
15. Markatou M, Basu A, Lindsay BG. Weighted likelihood equations with bootstrap root search. *Journal of the American Statistical Association*. 1998; 93:740–750.
16. McLachlan, GJ.; Peel, D. *Finite mixture models*. Wiley; New York: 2000.
17. Scott DW. Parametric statistical modeling by minimum integrated square error. *Technometrics*. 2001; 43:274–285.

18. Steimer J-L, Mallet A, Golmard JL, Boisvieux JF. Alternative approaches to estimation of population pharmacokinetic parameters: Comparison with the nonlinear mixed effects model. *Drug Metabolism Reviews*. 1984; 15:265–292. [PubMed: 6745083]
19. Thompson TJ, Smith PJ, Boyle JP. Finite mixture models with concomitant information: Assessing diagnostic criteria for diabetes. *Journal of the Royal Statistical Society, Series C: Applied Statistics*. 1998; 47:393–404.
20. Verbeke, G.; Molenberg, G. *Linear Mixed Models for Longitudinal Data*. Springer; New York: 2000.
21. Woodward WA, Parr WC, Schucany WR, Lindsey H. A comparison of minimum distance and maximum likelihood estimation of a mixture proportion. *Journal of the American Statistical Association*. 1984; 79:590–598.
22. Yau KKW, Lee AH, Ng ASK. Finite mixture regression model with random effects: Application to neonatal hospital length of stay. *Computational Statistics and Data Analysis*. 2003; 41:359–366.
23. Yeap BY, Catalano PJ, Ryan LM, Davidian M. Robust two-stage approach to repeated measurements analysis of chronic ozone exposure in rats. *Journal of Agricultural, Biological, and Environmental Statistics*. 2003; 8:438–454.
24. Zhan T, Chervoneva I, Iglewicz B. Generalized weighted likelihood density estimators with application to finite mixture of exponential family distributions. *Computational Statistics and Data Analysis*. 2011; 55:457–465. [PubMed: 20835375]
25. Zhang G, Ezura Y, Chervoneva I, Robinson PS, Beason DP, Carine ET, Soslowky LJ, Iozzo RV, Birk DE. Decorin regulates assembly of collagen fibrils and acquisition of biomechanical properties during tendon development. *Journal of Cellular Biochemistry*. 2006; 98:1436–1449. [PubMed: 16518859]

Figure 1. Histograms and kernel density estimates (Gaussian kernel) for conditional distributions in sample microscopic fields.

Figure 2.

Empirical relative % bias in estimates of finite mixture component means (MEAN1 and MEAN2) and log transformed standard deviations (LOG.SD1 and LOG.SD2). For each parameter, the first boxplot corresponds to ML estimates, the second one corresponds to MDPD estimates, and the third one corresponds to WLEE estimates.

Figure 3. Microscopic field specific estimates of means and log transformed standard deviations of 2-component normal mixtures estimated in stage 1 model (red circles for MLE, green diamonds for MDPD, purple triangles for WLEE).

Figure 4.

Correlation among microscopic field specific estimates of means and log transformed standard deviations of 2-component normal mixtures estimated in stage 1 model (red circles for MLE, green diamonds for MDPD, purple triangles for WLEE) by animal (animals 1–6 are decorin deficient and animals 7–12 are wild type).

Table 1

Fixed effects estimates and test results from stage 2 models

Parameter	Wild Type Estimate (SE)	Decorin Deficient Estimate (SE)	Difference Est. Diff. (SE)	P-value
ML estimation in Stage 1				
$\ln\sigma_1$	1.52 (0.11)	1.58 (0.12)	0.07 (0.05)	0.156
$\ln\sigma_2$	0.74 (0.19)	0.80 (0.22)	0.06 (0.06)	0.357
μ_1	34.71 (0.66)	37.24 (0.66)	2.53 (0.93)	0.010
μ_2	49.59 (1.44)	53.55 (1.44)	3.97 (2.04)	0.065
MDPD estimation in Stage 1				
$\ln\sigma_1$	1.45 (0.12)	1.51 (0.13)	0.06 (0.05)	0.243
$\ln\sigma_2$	0.86 (0.18)	0.95 (0.20)	0.09 (0.06)	0.149
μ_1	34.62 (0.75)	36.91 (0.75)	2.30 (1.06)	0.046
μ_2	49.69 (1.19)	53.47 (1.19)	3.78 (1.68)	0.029
WLEE estimation in Stage 1				
$\ln\sigma_1$	1.39 (0.11)	1.44 (0.13)	0.05 (0.05)	0.234
$\ln\sigma_2$	0.81 (0.17)	0.89 (0.19)	0.08 (0.05)	0.133
μ_1	34.72 (0.68)	37.07 (0.68)	2.35 (0.96)	0.022
μ_2	49.58 (1.29)	53.35 (1.29)	3.77 (1.82)	0.047

Table 2

Covariance parameter estimates from stage 2 models

Covariance Parameter	ML in stage 1 Estimate (SE)	MDPD in stage 1 Estimate (SE)	WLEE in stage 1 Estimate (SE)
Variance components of animal random effects			
$\tau^2_{\ln\sigma_1}$	0	0	0
$\tau^2_{\ln\sigma_2}$	0.0054 (0.0046)	0.0055 (0.0038)	0.0039 (0.0031)
$\tau^2_{\mu_1}$	0.15 (0.44)	0.86 (0.86)	0.39 (0.58)
$\tau^2_{\mu_2}$	7.09 (3.81)	1.74 (2.01)	4.06 (2.67)
Covariance of the random effects of microscopic field			
$\Sigma_{11}=\text{Var}(\ln\sigma_1)$	0.032 (0.006)	0.033 (0.006)	0.029 (0.005)
$\Sigma_{22}=\text{Var}(\ln\sigma_2)$	0.018 (0.005)	0.016 (0.004)	0.016 (0.004)
$\Sigma_{33}=\text{Var}(\mu_1)$	12.33 (2.34)	12.51 (2.76)	11.84 (2.34)
$\Sigma_{44}=\text{Var}(\mu_2)$	27.07 (5.51)	33.74 (7.9)	29.38 (6.21)
$\text{Corr}(\ln\sigma_1, \ln\sigma_2)$	0.51 (0.12)	0.79 (0.06)	0.77 (0.06)
$\text{Corr}(\ln\sigma_1, \mu_1)$	0.81 (0.05)	0.82 (0.05)	0.80 (0.05)
$\text{Corr}(\ln\sigma_2, \mu_1)$	0.32 (0.14)	0.57 (0.10)	0.52 (0.11)
$\text{Corr}(\mu_2, \ln\sigma_1)$	0.79 (0.05)	0.89 (0.03)	0.85 (0.04)
$\text{Corr}(\mu_2, \ln\sigma_2)$	0.34 (0.14)	0.68 (0.09)	0.58 (0.10)
$\text{Corr}(\mu_2, \mu_2)$	0.93 (0.02)	0.94 (0.02)	0.94 (0.02)