INVESTIGATION OF VARIANCE COMPONENTS IN THE MEDICAL EXPENDITURE PANEL SURVEY

Robert M. Baskin

Agency for Healthcare Research and Quality
540 Gaither Road, Rockville, MD 20850
rbaskin@ahrq.gov

Introduction

Statistical methodology has been developed to decompose total variance into sources of variation. These sources of variation are usually referred to as components of variance. There are different reasons for attempting to produce variance components. In models decomposing the variance is often done in order to increase the accuracy of the model. In the finite population setting, variance is usually decomposed into variance components to explain variation due to different levels of sampling.

Variance components are often of secondary importance in any analysis, but if the variance components are of interest as a focus of an analysis, then for theoretical reasons, Bayesian methodology can be argued to be the best method available for estimation. Although Bayesian methods have a theoretical advantage in some respect, there is not always a reliable practical method of estimation available to the Bayesian statistician. However, the advent of the Markov Chain Monte Carlo (MCMC) technology and advances in software in recent years, have made Bayesian methods a more practical alternative in many settings.

Some prior investigations into application of these MCMC methods to estimation of variance components for survey data reported only partial success. The primary purpose of this investigation is to determine if MCMC methodology is practical for estimating variance components for two analytical variables of interest in the Medical Expenditure Panel Survey (MEPS). Evaluation of this objective will be based on convergence criteria developed in support of the MCMC methodology.

Background: MEPS Survey and Sample Design

The MEPS is a complex national probability sample survey sponsored by the Agency for Healthcare Research and Quality (AHRQ). MEPS is designed to provide nationally representative estimates of health care use, expenditures, sources of payment, and insurance coverage for the U.S. civilian noninstitutionalized population. MEPS consists of a family of three interrelated surveys with the Household Component (HC) as the core survey. Healthcare expenditures is considered one of the primary analysis variables in MEPS and is published by year. A self reported measure of health status is collected and published by round. In 2002 MEPS-HC began collecting SF-12 which is a proprietary composite index of health status. These three variables are used to investigate estimation of hierarchical variance components.

The sample of households for the MEPS-HC is a subsample of households that responded to the prior year's National Health Interview Survey (NHIS), conducted by the National Center for Health Statistics. The NHIS sample design is a multi-stage cluster design with unequal probability of selection. As stated in an NCHS Methods Report, Botman SL, et al (2000), "The first stage consists of a sample of 358 primary sampling units (PSUs) drawn from approximately 1,900 geographically defined PSUs that cover the 50 States and the District of Columbia. A PSU consists of a county, a small group of contiguous counties, or a metropolitan statistical area."

Within a PSU there are two types of second-stage units. The first type is area segments that are defined geographically from 1990 Census information and contain an expected 8 or 12 addresses. The second type is permit area segments that cover geographical areas containing housing units built after the 1990 census. The permit area segments are defined using updated lists of building permits issued in the PSU since 1990 and contain an expected four addresses. Within each segment all occupied households at the sample addresses are targeted for interview.

The total NHIS sample of PSUs is subdivided into four separate panels and each panel is a representative sample of the U.S.

population. The MEPS-HC selects a subsample of households from two of the four NHIS panels and follows households that were interviewed in NHIS. The MEPS-HC uses an overlapping panel design in which data are collected through a series of five rounds of interviews over a two and one-half year period. Details on the weights and the MEPS sample design can be found in (Cohen, 1997; Cohen, 2000).

Variables of Interest

In MEPS, a self-reported measure of health status is also collected and reported by round of collection. It is measured on a five point Likert scale with a score of one corresponding to excellent health and a score of five representing poor health. Other measures of health status are collected in later panels. As stated in the MEPS 2002 full year file documentation, "The Self Administered Questionnaire contained two measures of health status, the Short-Form 12 (SF-12 (r), a registered trademark) and the EuroQol 5-D (EQ-5D). These are two of the more widely used measures of health status." A technical discussion of these two measures and their use in cost-utility analysis can be found in Lawrence and Fleishman (2002).

Estimation of the Variance Components for the MEPS Sample

In survey sampling, the most common use of variance components is for sample allocation across the stages of sampling. There are several references on variance components in survey settings. Wolter (1985) and Korn and Graubard (2003) provide useful references. In the case of the MEPS-HC, this kind of allocation is not a possibility because MEPS-HC is a subsample of the existing NHIS sample. It should also be pointed out that indicators of segment membership are not available on the MEPS-HC datasets. Thus, even if at some point if it were desired to allocate a totally new sample based on the current one, it would be impossible to do so without the segment level indicators.

However, the variance can be decomposed in various ways such as into PSU level and below PSU level components based on the information that is available. Even though this can be done, it would be an unusual application that would require this information. As it turns out, for the case of health status, two unusual applications did arise.

The first application involved *predicting* health utility from health status. The technical differences between a health utility index and a health status score are beyond the scope of this work but can be found in Lawrence and Fleishman (2004). In that work, *predicting* a utility index from health status was investigated. The reason for building a prediction model as opposed to fitting a model was based on the fact that many datasets collect health status but do not collect a utility index. If the MEPS-HC data could be used to build a prediction model of health utility based on health status, then in other datasets with sufficient representation of specific diseases but no utility index, the prediction model could be used to predict the impact of specific disease on health utility. This idea is explored in Lawrence and Fleishman (2004).

Formulas for prediction intervals are given, for example, in Miller (1981). The formula in Miller was tested using two years of MEPS data. Data from 2001 MEPS-HC was used to build a predictive model. Then the prediction was applied to the 2002 MEPS-HC. The distribution of the prediction was compared to the distribution of the actual utility value available in the 2002 dataset. In terms of the mean, the predicted values were almost the same as the actual measured index values in the 2002 dataset. If the predicted values were treated as another form of observed values and the prediction part of the variance ignored, then the variance of the predicted values underestimated the actual values by about twenty five percent. An application of the formula in Miller estimated the variance of the predicted values to be about 90% of the actual variance of the utility values. To try to explain this discrepancy an analysis of the variance components of health status was undertaken. These components are estimated below.

The second application involves a different decomposition of the variance than the first. Pfeffermann et. al. (1998) describe the Probability Weighted Iterative Generalized Least Squares method for fitting multilevel models to complex survey data. This requires the fitting of variance components which are along the lines of the traditional concept of variance components. These are estimated for both health status and for the general health component of SF-12.

The estimation of components of variance for expenditures, which is a primary variable of investigation in MEPS-HC, is an area for future research.

Marginal versus Hierarchical Models

First consider a general linear model. The general linear model can give rise to two views of the model that are used to

estimate variance components.

Assume that **Y**ij denotes a vector of variables of interest for subject i in PSU j. In the case of health status it is a five component vector representing the five rounds of measurement. In this case the general linear model is represented as $Y_{ij} = X_i\beta + Z_i b_i + \varepsilon_{ij}$, where, in frequentists' language, β represents the fixed component and b_i represents the random component. The matrices X and Z are assumed to be known. The b_i represent a set of subject specific regression coefficients. In repeated measures the b_i describe how the trajectory of the i^{th} individual deviates from the population. The b_i have mean 0 and variance matrix D. The ε represent deviation for each measure and have mean 0 and variance Σ. It is often stated that the b and the ε are "assumed" to be independent but in fact a model can be constructed as such using Hoeffding projection.

For simplicity of presentation assume that the random terms are normally distributed. A fully hierarchical model is specified by $Y_{ij} \mid b_i \sim N(X_i\beta, +Z_i b_i, \Sigma_i), b_i \sim N(0, D)$ whereas the marginal model is given by $Y_{ij} \sim N(X_i\beta, Z_i DZ' + \Sigma_i)$. These two models appear to be equivalent and are often assumed to be so.

Molenbergs and Verbeke (2004) give an accessible explanation of the important differences between the hierarchical model and the marginal model, and there is more detail provided in Verbeke and Molenbergs (2000). The differences between the two are exactly in the assumptions and constraints on the variance components. Specifically, for the hierarchical model, the matrices D and Σ are both positive definite, while in the marginal model the variance matrix must be positive definite but there is no restriction on the individual matrices D or Σ.

Nelder (1954), under the simplest setup, gave the first published explanation, in the language of split plot designs, of the conditions necessary to make the marginal and hierarchical models equivalent. The variance matrix in Nelder's article was of the form given by $\tau^2 J + \sigma^2 I$ where J is a square matrix with all entries equal to one. In this setting Nelder derived the constraint for τ^2 to be positive. The interpretation of τ^2 if it is negative is that it is actually a covariance.

Note that this constraint is necessary for the population parameter τ^2 to be positive and is not referring to a negative estimate of τ^2. The problem of negative estimates of positive population variance components is discussed in section 3.4 of Searle, Casella, and McCulloch (1992). But the problem discussed in Nelder (1954) is that the underlying population parameter from the marginal model is actually negative.

Why Hierarchical and Why Bayesian

Determining the constraints necessary to make the variance components positive can be very difficult or impossible under the marginal model. However, under the hierarchical model the constraints for positive variance components follows from the fact that the variance matrices must be positive definite at each level of the model. Furthermore, hierarchical modelling is a natural way to model a multi-stage sample. For these reasons a hierarchical model is used to estimate the variance components.

Because the desire is to estimate the hierarchical variance components, a Bayesian method is used. As stated in Molenberghs and Verbeke (2004), "Arguably, a satisfactory treatment of the hierarchical model is only possible in the Bayesian setting".

Finite Population versus Infinite Population Bayesian Estimates

The estimates produced here can be considered valid from an infinite population perspective. The theory for this is sufficiently developed in, for example, Bernardo and Smith (2000). From the perspective of finite population Bayesian methods a reference to the theory can be found for example in Ghosh and Meeden (1997). Arguably however, there are gaps in the theory that need to be filled in order to produce Bayesian estimates of variance components in a finite population setting. There are indications in Chapter Six of Ghosh and Meeden that this approach is a reasonable area of future research.

Method of Estimation

Previously Baskin (1993) and Baskin and Johnson (1995) investigated the use of Bayesian methodology to estimate variance components of a complex survey. The methods used in both those papers were based on MCMC. In the first paper, the

MCMC methodology worked very well to produce estimates of variance components but in the second paper the MCMC methodology was reported as definitely not convergent.

There are now newer interfaces to BUGS software which provide ease of use and reliability of code for MCMC. The BUGS code has been successfully running for over a decade now. This is the code basis for the current evaluation.

For the current data, the hierarchical models were fit using the open source version of WinBUGS, which is directly callable in R, from the R Development Core Team (2005), through the package BRugs. Because the models were fit in BUGS, the method of estimation used was an MCMC method. For each model fit, initial values are necessary to start the MCMC method. For the models fit, initial values for the parameters were set equal to a method of moments estimates of parameter plus a random number. Three independent chains were started for each model, so three independent random numbers were added to the method of moments values to produce three independent initial starting values. Starting independent chains allows assessment of sensitivity of the final estimate of the posterior to initial values.

For all hyperparameters in the model associated with means or regression coefficients, noninformative conjugate priors were used. For the variance components, an inverse Gamma distribution was the form of the prior distribution, but the hyperparameters were chosen in a way that made this distribution nearly constant over the entire support of the variance component. This is a typical approach to modeling variance components in BUGS.

Because of the size of the dataset and the running time required, all of the models were built only for subjects who were in Census Region 1 in round 1. This reduced the size of all models by about one fourth to approximately 2,000 subjects.

The model for the self reported health status models the mean for the individual as a random normal deviate around the PSU mean plus a random individual level trajectory over the five rounds of collection. Each PSU level mean was generated as a random normal deviate around the strata level mean.

The model for the SF12 models the mean for the individual as a random normal deviate around the PSU. There is only round of collection for SF12 in a panel so there is no trajectory over time. Each PSU level mean was generated as a random normal deviate around the strata level mean. The code for the expenditure model is in the appendix.

Results

The model for health status was run for a burn-in of 20,000 samples and then monitored for 12,000 samples. The convergence criteria indicated that convergence was achieved. Displaying the results for the subject level parameters is cumbersome since there are 1,740 parameters to describe. Visual inspection of the subject level parameters indicated no problems in the final density, in the sample history, or in the convergence measures such as the Brooks-Gelman-Rubin (BGR) statistic.

The two levels of the variance components investigated for this variable were denoted sigma[1] for the PSU level component and sigma[2] for the lower level component. The model showed clear signs of convergence for these parameters and the estimated posterior densities are shown below. The standard BUGS output statistics from the posterior density of these components is in Table 1.

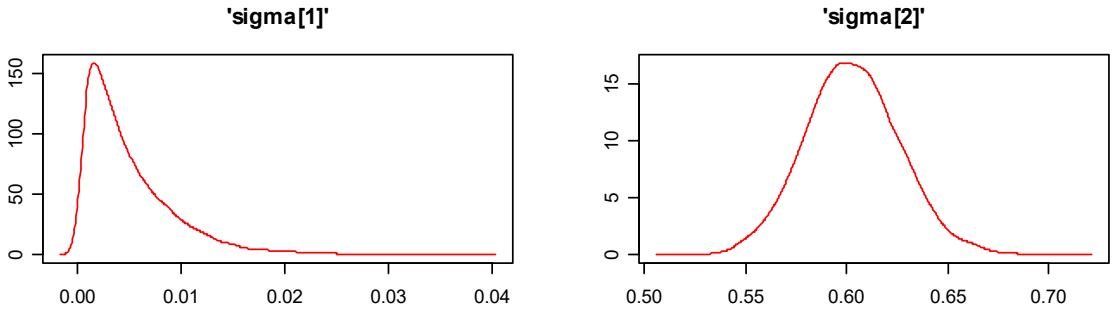
Table I.

Distributional information from the densities of the components

	Mean	Sd	MC error	2.50%	Median	97.50%
sigma[1]	0.005036	0.004345	0.0001853	0.0005541	0.003727	0.01662
sigma[2]	0.6027	0.02338	0.0002266	0.5578	0.6022	0.6502

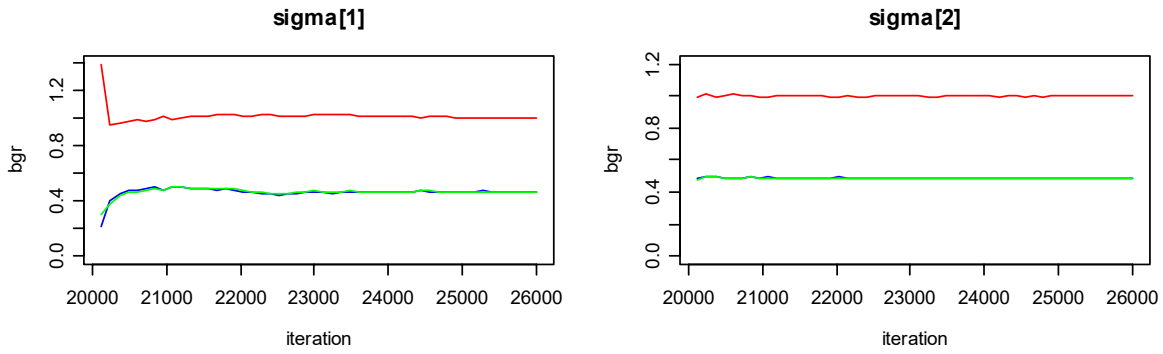
The posterior densities for the variance components are presented in Figure 1.

Figure I.
Densities of the components



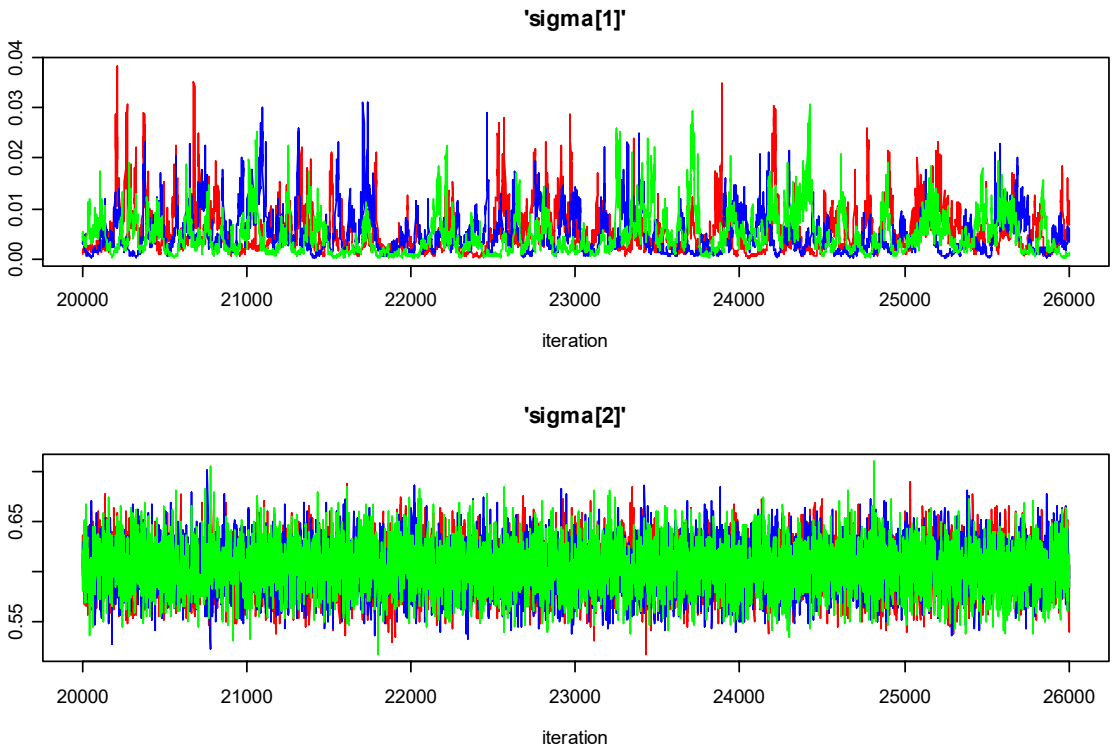
The Brooks-Gelman-Rubin Statistic is a graphical measure to assess convergence. In order to claim the model has achieved convergence, the upper line should converge to 1.00 and the lower two lines should converge to 0.40. References for this idea can be found in the manual for WinBUGS available through the BRugs package. For the variance components of Self Reported Health Status (SRHS) the graph is stable after 20,000 iterations, as shown in Figure 2.

Figure 2.
Graphs of the Brooks-Gelman-Rubin statistic for components of SRHS



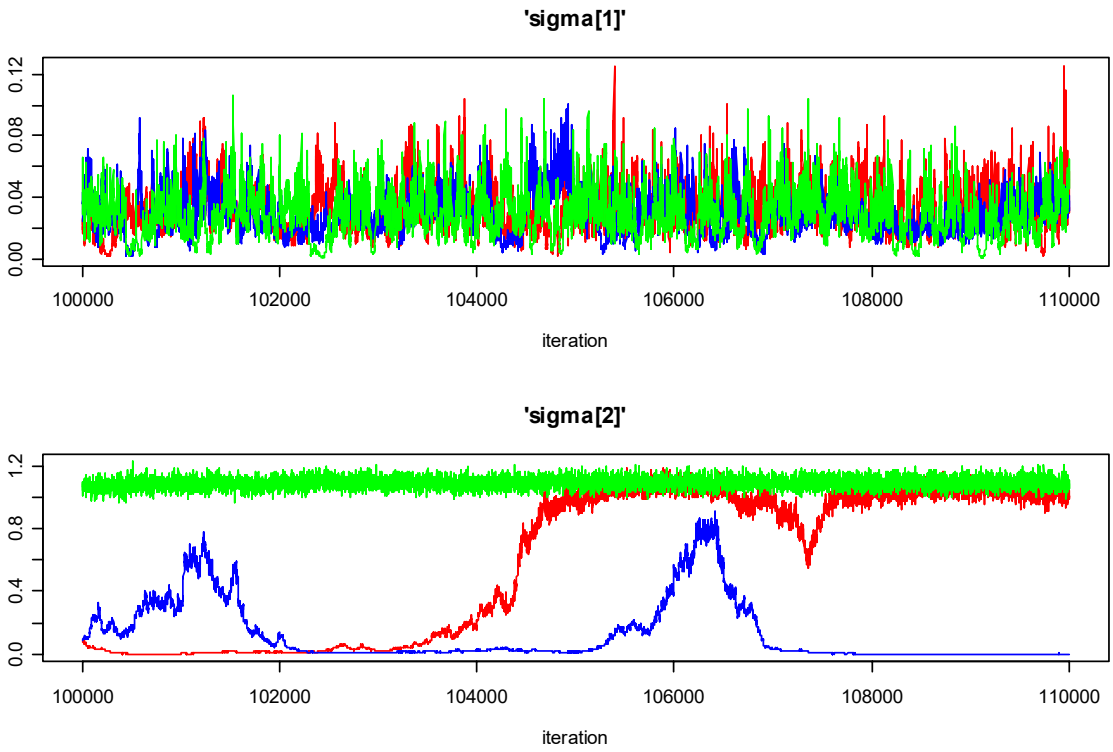
In order to judge convergence the sample histories for the three independent chains should be graphed. The histories for the independent chains for the variance components of Self Reported Health Status are presented in Figure 3.

Figure 3.
Sample Histories for Self Reported Health Status components



The variable used is a component of the SF12 for general health status and is very similar to the self reported health status. However, the SF12 questions are not asked of all respondents so there are many structural missing values for this component of SF12. The model for SF12 initially appeared to converge but further runs indicated that the below PSU variance component had independent chains diverging to two distinct values. This indicates that the current model cannot be used. This model was further allowed to run through a million iterations and the same problem continued to appear. This is quite counterintuitive since the SF-12 variable is very similar in both distribution and underlying measurement concept to the Self Reported Health Status variable for which the model converged quickly. The model structure is not complicated for this variable so there is little reason to believe that this variable should have a problem. There is one aspect of the component of SF-12 variable that is different and that is that it has more missing values than Self Reported Health Status. The missingness for this variable is a form of structured missingness due to the fact that not all respondents are asked the input questions to construct the SF-12 index. The sample histories of the components for SF-12 are in Figure 4. It is observed that the three independent chains for the second component of SF-12 show a striking problem of non-convergence.

Figure 4.
Histories of the components for SF-12



Conclusion

This research project examined selected aspects of estimating different decompositions of the variance associated with health status and medical expenditures in a setting of multi-stage sampling. The models for health status fit well and converged quickly. The models for SF-12 clearly has convergence problems. This answers the primary objective of the research in that the MCMC methodology for estimating variance components can only be claimed to be partially successful. In a finite population setting, the Bayesian theory and methodology need to be developed further.

References

Baskin, Robert M. (1993), ``Estimation of variance components for the U.S. Consumer Price Index via Gibbs sampling", *ASA Proceedings of the Section on Survey Research Methods*, 808-813

Baskin, Robert M. , and Johnson, William H. (1995), ``Estimation of variance components for the U.S. Consumer Price Index", *ASA Proceedings of the Section on Survey Research Methods*, 126-131

Bernardo, J. and Smith, A. (2000), "Bayesian Theory", John Wiley & Sons (New York; Chichester)

Botman SL, Moore TF, Moriarity CL, and Parsons VL. Design and estimation for the National Health Interview Survey, 1995–2004. National Center for Health Statistics. Vital Health Stat 2(130). 2000.

Ghosh, Malay and Meeden, Glen (1997) "Bayesian Methods for Finite Population Sampling", Chapman Hall(New York).

Korn, E. and Graubard, B. (2003) "Estimating Variance Components by Using Survey Data", Journal of the Royal Statistical Society, Series B, Methodological, 65 , 177-190

Lawrence, W.F. and Fleishman, J.A. (2004) "Predicting EuroQoL EQ-5D Preference Scores from the SF-12 Health Survey in a Nationally Representative Sample" Medical Decision Making, Vol. 24, No. 2, 160-169 (2004).

Miller, Rupert G. (1981), ``Simultaneous statistical inference", Springer-Verlag Inc (Berlin; New York).

Molenberghs, Geert , and Verbeke , Geert (2004), ``Meaningful statistical model formulations for repeated measures", Statistica Sinica, Vol. 14, No. 3, 989-1020 (2004).

Nelder, J.A.(1954), "The Interpretation of Negative Components of Variance", Biometrika, 41, 544-548.

Pfeffermann, D. , Skinner, C. J. , Holmes, D. J. , Goldstein, H. , and Rasbash, J. (1998), ``Weighting for unequal selection probabilities in multilevel models (Disc: p41-56)", Journal of the Royal Statistical Society, Series B, Methodological, 60 , 23-40

R Development Core Team (2005). "R: A language and environment for statistical computing." R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Searle, S. R. , Casella, George , and McCulloch, Charles E. (1992), ``Variance components", John Wiley & Sons (New York; Chichester)

Verbeke, Geert , and Molenberghs, Geert (2000), ``Linear mixed models for longitudinal data", Springer-Verlag Inc (Berlin; New York)

Wolter, Kirk (1985), "Introduction to Variance Estimation", Springer-Verlag Inc (Berlin; New York)

Zhou, X. and Gao, S. (1997) "Confidence Intervals for the Log-Normal Mean", Statistics in Medicine, 16, 783-790.