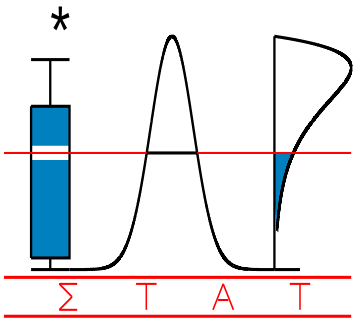
T E C H N I C A L

R E P O R T

0615

MODELLING ETHANOL-INDUCED SLEEPING TIME IN INBRED STRAINS OF MICE: A REPEATED MEASURES DESIGN WITH LEFT-CENSORED DATA.

R. BRAEKERS and P. MARKEL



I A P S T A T I S T I C S

N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

Modelling ethanol-induced sleeping time in inbred strains of mice : A repeated measures design with left-censored data.

Roel Braekers
Hasselt University, Center for Statistics, Belgium

and

Paul Markel
Department of Addiction Studies, Psychology & Social Work,
Minot State University, Minot (ND), U.S.

Abstract

In this paper we present a repeated measures analysis of ethanol-induced anesthesia (sleeping time) in mice from the inbred long-sleep (ILS) and short-sleep (ISS) strains of mice and their derived F_1 and F_2 generations. Due to the difficulty in the assessment of ethanol-induced sleeping time, some of the measurements are left-censored by a fixed detection limit of 1 min. Furthermore we note that some of the animals in the F_2 population are related since we are able to identify the different litters of mice in this population. Unlike previous analyses of this data set in Markel and Corley (1994), Markel et al. (1995a and 1995b), we develop a method that is able to cope with the full complex structure of the ethanol-induced anesthesia data set. Therefore we extent the linear mixed models framework to handle left-censored observations. As results, we first determine which genetic factors have an significant influence on the ethanol-induced sleeping time. Afterwards we also consider the influence of different environmental factors on this trait. In each situation we give estimates for the heritability and the number of effective genes for this trait.

Keywords: detection limit, genetic analysis, heritability, left censoring, linear mixed models.

1 Introduction

In this article we analyze a repeated measurements design experiment on ethanol-induced sleeping time in laboratory mice given by Markel and Corley (1994), Markel et al. (1995a and 1995b). Like in many studies with psychological phenomena, we see that some mice in this study did not fall asleep from the dose of ethanol used in the experiment or had a sleeping time which was shorter then the time considered in the design to assess whether a mouse has gone to sleep after an injection. In this way we are confronted with observations which cannot be accurately measured below a fixed value. We call these observations left-censored. In the previous articles the censoring aspect in this data set was either neglected or treated by using Cohen corrected estimates for the mean and variances (Cohen (1959)).

We propose in this article a maximum likelihood method which on the one hand can handle the left-censoring in this data set and on the other hand is able to cope with the repeated measurements design and the dependence between the measurements of animals from the same litter. Our method is an extension of the linear mixed models framework to handle left-censored observations and provides a more flexible and natural tool to express the idea's given in Markel et al. (1995b). Thiébaut and Jacqmin-Gadda (2004) considered a similar method in their article on longitudinal left-censored repeated measures however they were not able to handle nested random effects as we will do. As in the previous articles on the ethanol-induced anesthesia data set, we focus in this paper on the determination of both genetic and environmental factors for the ethanol-induced sleeping time. Furthermore we find an estimate for the heritability of this trait and for the number of effective genes which determine this trait.

This article is build up as follows. In Section 2 we introduce the ethanol-induced anesthesia data set in full detail. We discuss in Section 3 the mathematical methodology of our maximum likelihood method. Afterwards in Section 4 we apply this methodology on the data set and give the results of the performed analyses. In Section 5 we finish this article by a conclusion.

2 Ethanol-induced anesthesia data set

The ethanol-induced anesthesia data set was created between 1991 and 1993, and contains repeated measurements on ethanol-induced sleeping time for four different populations of

mice, divided over three generations. The first generation consists of two inbred strains, ILS and ISS, which were created by sib mating and selection for length of sleeping time from standard laboratory mice ($\pm$ 30 generations). In the ILS strain, the mice have a long sleeping time of more than 3 h, while in the ISS strain the mice have a short sleeping time of less than 10 min. These strains were crossed, in the way that ILS females were mated with ISS males and vice versa, to produce both reciprocal F_1 strains, L/S and S/L respectively. The F_1 strains were afterwards crossed in all combinations to produce four segregating F_2 strains $L/S \times L/S$, $L/S \times S/L$, $S/L \times L/S$ and $S/L \times S/L$.

All mice were produced in a specific pathogen-free (SPF) colony and were weaned at 25 days of age. At approximately 50 days they were removed to a non-SPF environment where they remained for 1 week before testing. The mice were maintained on a 12 h light cycle (07.00 to 19.00) and were allowed food (Wayne diet by Teklad) and water ad lib.

In this article, we look at the ethanol-induced sleeping time and how this depends on genetic and/or environmental factors. The design of this study is a repeated measurements setting. The mice were initially injected at 55-65 days of age (Trial 1) and again 7-10 days later (Trial 2). In the afternoon prior to the day of testing, the animals were weighted and moved into the testing room, where they remained until completion of a trial. The mice were injected intraperitoneally with a 4.1 g/kg dose of ethanol (20% w/v in saline) between 09.00 and 13.00. The order of individual injections was retained from the first trial to the next. After an injection the animal was placed on its back in a plexiglass trough and was considered to be anesthetized if it did not turn over more than three times within the first minute. Repeated attempts to observe this behavior were made within 15 min after the injection. Anesthetic recovery was indicated when an individual turned over three times within 1 min after being anesthetized. The sleeping time of a mouse was measured as the time interval between observed anesthesia and the final minute of recovery. We note that in the assessment of sleeping time, the recording is left-censored by a fixed detection limit of 1 min. For a mouse in which repeated attempts to place it on its back failed, we know that the animal probably did not fall asleep or had a sleeping time of less than 1 min. We consider the measurements of these mice as left-censored observations. Furthermore we allowed all mice to sleep until recovery.

Next to the ethanol-induced sleeping time, we recorded in each animal several covariates. In this article we will focus on only a few of those, like sex, coat color, cross, mating pair, weight, birthday and trial days for trial 1 and trial 2. Using the variables cross, mating pair and birthday, we are able to identify in each population the different litters. The family link between animals of the same litter allows us to get an idea about the

heritability of sleeping time in mice. The variable coat color of a mouse is dichotomized in the analysis and will be used to study whether an albino mouse reacts differently to alcohol than a non-albino mouse. Markel and Corley (1994) found that the gene coding for albinism (Tyr) had an effect on the ethanol-induced sleeping time. They posed that either this gene or a gene closely linked to it, is important for sleeping time.

3 Methodology

In this section we develop a model to analyze the ethanol-induced sleeping time in laboratory mice. As we saw in the previous section, the ethanol-induced anesthesia data set has a complex structure. On the one hand we see that some of the observations in this data set are left-censored, and on the other hand we have to take into account that this data set has a repeated measurements design in which each mouse is measured twice in a period of 10 days. For the different populations we are able to identify the litter of each animal. However, only for the segregating F_2 population will this provide extra information in the model, as we show further on.

In the development of a model for the sleeping time, we divide this problem in two sub-problems. As first sub-problem we handle the left-censoring in this data set. We consider a latent variable Y which represents the ethanol-induced sleeping time in a mouse. However due to the detection limit of 1 min in the assessment of this sleeping time, the observed sleeping time is given by a censored variable $Z = \max(Y, 1)$ which is the maximum of the underlying sleeping time and the detection limit. For each measurement we define a censoring-indicator $\delta = I(Y > 1)$ which indicates whether the mouse has slept. This construction allows to model the underlying ethanol-induced sleeping time Y without the problems in the practical assessment of this time.

As second sub-problem in the analysis of the ethanol-induced anesthesia data set, we consider the complex structure in the different populations of mice. To handle this sub-problem, we define models within the linear mixed models framework for the underlying ethanol-induced sleeping time. For the isogenic populations ISS, ILS and F_1, we take the following linear mixed model. If we denote by Y_{ij}^p the ethanol-induced sleeping time for animal i at trial j in the population $p \in \{ILS,\ ISS,\ F_1\}$, we define the model

$$Y_{ij}^p = \mu^p + \tau_i^p + \varepsilon_{ij}^p, \quad j = 1, 2,\ i = 1, \ldots, n_p$$

with $\tau_i^p \sim \mathrm{N}(0, \sigma_p^2)$, $\varepsilon_{ij}^p \sim \mathrm{N}(0, \varepsilon_p^2)$ and τ_i^p, ε_{ij}^p independent.

Note that we assume a two-level hierarchical model for the isogenic populations ILS, ISS and F_1. On the lowest level we have the individual measurements while at the second level we have the different animals in the population. Since all the animals in these populations are genetically identical, animals from the same litter do not provide an extra source of variability in this model. In the segregating F_2 population, the different animals of a litter are not genetically identical and provide a way to estimate the heritability of the ethanol-induced sleeping time. Therefore we consider in this population a three-level hierarchical model with on the lowest level the individual measurements for the sleeping time. On the second level we have the different animals and on the third level we have the different litters. If we denote by $Y_{ijk}^{F_2}$ the sleeping time for animal i from litter k at trial j, we define the model

$$Y_{ijk}^{F_2} = \mu^{F_2} + \tau_k + \tau_{ik} + \varepsilon_{ijk}, \quad j = 1, 2,\ i = 1, \ldots, n_k,\ k = 1, \ldots, n_{F_2}$$

with $\tau_k \sim \mathrm{N}(0, \tau^2)$, $\tau_{ik} \sim \mathrm{N}(0, \sigma_{F_2}^2)$ and $\varepsilon_{ijk} \sim \mathrm{N}(0, \varepsilon_{F_2}^2)$. We assume that τ_k, τ_{ik}, ε_{ijk} are independent.

From the design of this study we know that the different mouse populations are genetically related. In the remaining of this section we describe how the parameters in the different linear mixed models are linked and how we find a natural interpretation for these parameters by connecting them to quantities commonly used by geneticists. Furthermore we define estimators for the heritability of ethanol-induced sleeping time and for the number of genes which determine this trait in mice.

We know from the literature that a phenotypic variance V_P of a trait can be split in three components, a genetic component V_G, an environmental component V_E and an interaction between these components V_{GE},

$$V_P = V_G + V_E + V_{GE}.$$

In this article we do not consider this interaction term since all the animals in this study were bred and tested in the same environment and under the same conditions. We split the environmental variance V_E further in a common environmental variance V_{Ec} and a specific environmental variance V_{Es},

$$V_E = V_{Ec} + V_{Es}.$$

The common environmental variance expresses in this study the variability in ethanol-induced sleeping time due to environmental factors which do not change between the two trial sessions, while the specific environmental variance is the variability which results from changed environmental factors. We note for each isogenic population that the

observed variance of the ethanol-induced sleeping time in the linear mixed model is given by

$$\mathrm{Var}(Y_{ij}^p) = \sigma_p^2 + \varepsilon_p^2, \quad p \in \{ILS, ISS, F_1\}.$$

Since there is no genetic variability in the animals of such a population, we see that this phenotypic variance is equal to the environmental component

$$V_P^p = V_E^p = V_{Ec}^p + V_{Es}^p = \mathrm{Var}(Y_{ij}^p) = \sigma_p^2 + \varepsilon_p^2$$

From the linear mixed model we know that the variance σ_p^2 expresses the variability between the different animals and the variance ε_p^2 gives the variability between the different measurements within an animal. Therefore we see that we can identify the different terms at both sides of the previous equation, as

$$V_{Ec}^p := \sigma_p^2 \quad \text{and} \quad V_{Es}^p := \varepsilon_p^2.$$

We note that in this analysis, each isogenic population is allowed to have a different value for the phenotypic variance. This was not possible in the method of Markel et al. (1995b).

Unlike for the isogenic populations, we know that the phenotypic variance in a segregating population is due to the genetic variance V_G. In this study we were able to identify the different litters in the F_2 population. For these animals, we know that they share a part of the same genetic variance, namely the part which is common between brothers and sisters in a sibship and which is given by $\frac{1}{2}V_A + \frac{1}{4}V_D$. V_A represents in this expression the additive variance and V_D is the dominance variance. In the linear mixed model for the segregating F_2 population, we find this quantity as the covariance between measurements of animals from the same litter. Using this idea, we find that for two animals i and i' of the same litter,

$$\mathrm{Cov}(Y_{i1k}^{F_2}, Y_{i'1k}^{F_2}) = \mathrm{Var}(\tau_k) = \tau^2 := \frac{1}{2}V_A + \frac{1}{4}V_D.$$

After we linked the different variances in the linear mixed models with genetic quantities, we can define the heritability of ethanol-induced sleeping time in mice. From the literature we know that the heritability is defined as the ratio V_A/V_P of the additive variance on the phenotypic variance. We note from the linear mixed models that we cannot estimate the additive variance V_A directly. However we can give an upper bound for the heritability. Based on the model for the F_2 population, we find that

$$h^2 = \frac{V_A}{V_P^{F_2}} \leq \frac{V_A + \frac{1}{2}V_D}{V_P^{F_2}} = \frac{2\tau^2}{\tau^2 + \sigma_{F_2}^2 + \varepsilon_{F_2}^2}.$$

Note that we only need variance terms of the segregating F_2 population to estimate the heritability, which is an improvement on the method used in Markel et al. (1995b). Next to the heritability, we also estimate the number of effective genes which regulate the ethanol-induced sleeping time. In this article we use the Castle-Wright estimator (Wright (1968)). This estimator compares the squared difference of the average sleeping time in the ILS and ISS populations with the additive variance V_A and is given by

$$n_{CW} = \frac{(\mu^{ILS} - \mu^{ISS})^2}{8V_A}.$$

As we have shown before, we cannot define the additive variance V_A as a combination of the variance-parameters from the linear mixed models. Therefore we derive a lower bound for the number of effective genes of ethanol-induced sleeping time, which is given by

$$n_L = \frac{(\mu^{ILS} - \mu^{ISS})^2}{8(V_A + \frac{1}{2}V_D)} = \frac{(\mu^{ILS} - \mu^{ISS})^2}{16\tau^2}.$$

Next to the variance parameters of the linear mixed models, we also link the mean parameters in these models to quantities commonly used in genetics. This allows to investigate the influence of several genetic factors on ethanol-induced sleeping time. Due to the construction of the different mouse populations in this study, we know from the literature that the average sleeping times in these populations are linked to each other and are given by

$$\begin{aligned} \mu^{ILS} &= m + a + i \\ \mu^{ISS} &= m - a + i \\ \mu^{F_1} &= m + d + l \\ \mu^{F_2} &= m + \frac{d}{2} + \frac{l}{4}. \end{aligned}$$

In these expressions the parameter m represents the common average sleeping time for the four populations. This parameter is equal to the intercept parameter in a linear mixed model. The parameter a is called the breeding value and gives the average additive deviation from the common average sleeping time. This parameter is only shown in populations with homozygous individuals and represents the average effect on sleeping time of the different genes which regulate this trait. The parameter d is called the dominance effect and stands for the average dominant deviation from the common average sleeping time m. This dominance effect represents the interaction between two different alleles of a gene at the same locus. As last parameters in the average sleeping time of

the four mouse populations, we have the parameters i and l. These parameters give the average interaction effects between two genes for sleeping time at different loci. The parameter i gives an interaction between homozygous loci while the parameter l gives an interaction between heterozygous loci (Mather and Jinks (1982)).

To estimate the different parameters of these linear mixed models, we consider a maximum likelihood method which is adapted for left-censored observations. In the construction of the likelihood function, we first consider the contribution of an observation from an isogenic population. Conditionally on the random effect for the repeated measurements within a mouse, we have that the contribution is given by

$$\prod_{j=1}^{2} f(Y_{ij}^{p} \mid \tau_i^{p})^{\delta_{ij}} F(Y_{ij}^{p} \mid \tau_i^{p})^{1-\delta_{ij}}.$$

The likelihood function for an isogenic population is found by integrating out the random effect.

$$L^{p} = \prod_{i=1}^{n_p} \int \prod_{j=1}^{2} f(Y_{ij}^{p} \mid \tau_i^{p})^{\delta_{ij}} F(Y_{ij}^{p} \mid \tau_i^{p})^{1-\delta_{ij}} f(\tau_i^{p}) d\tau_i^{p}.$$

with $p \in \{ILS,\ ISS,\ F_1\}$.

In the segregating population F_2, the contribution of a mouse, conditional on the random effects, is given by

$$\prod_{j=1}^{2} f(Y_{ijk}^{F_2} \mid \tau_k, \tau_{ik})^{\delta_{ijk}} F(Y_{ij}^{F_2} \mid \tau_k, \tau_{ik})^{1-\delta_{ijk}}.$$

The likelihood function for the segregating F_2 population is found by integrating over the random effects.

$$L^{F_2} = \prod_{k=1}^{n_{F_2}} \int \prod_{i=1}^{n_k} \int \prod_{j=1}^{2} f(Y_{ijk}^{F_2} \mid \tau_k, \tau_{ik})^{\delta_{ijk}} F(Y_{ijk}^{F_2} \mid \tau_k, \tau_{ik})^{1-\delta_{ijk}} f(\tau_{ik}) d\tau_{ik} f(\tau_k) d\tau_k.$$

In the design of the study, it is clear that the different mouse populations are genetically related. However this does not mean that some test animals in one population are the parents of test animals in another population. In this study the parental animals used to breed the F_1 and F_2 population, are not used as test animals. This implies that the test animals in different populations are independent. The overall likelihood function of this study is the product of the likelihood functions in each population separately.

4 Results

4.1 A descriptive data exploration

In this section we perform a descriptive data exploration of the ethanol-induced anesthesia data set to gain more insight into this data set and to verify whether the normal distribution is a reasonable distribution for the underlying ethanol-induced sleeping time. This normal distribution is a key assumption in the methodology of the linear mixed models, described in the previous section. From the literature on population genetics we have another motivation for this normal distribution. Each genetic trait is determined by, in general, several genes with different alleles at each locus. In genetic theory, it is assumed that each combination of alleles for the different genes gives a discrete value in the phenotype of this trait. Since the number of combinations is in most cases very large, the number of discrete phenotypic values is also large and therefore the normal distribution will be a good approximation of the true situation. This approximation is commonly used in genetics and is important for studies in which no genetic information is available to distinguish between the different allelic combinations, as is the case in this article (Falconer (1989), Mather and Jinks (1982)).

Due to the left-censoring by a fixed detection limit, we note that verifying the normality assumption is not straightforward. The observed data has under this assumption a censored normal distribution. If we want to verify whether the ethanol-induced sleeping time has a normal distribution, we work as follows. We assume that this sleeping time is normally distributed $X \sim N(\mu, \sigma^2)$ with unknown parameters μ and σ^2. For each animal we do not observe the ethanol-induced sleeping time directly but only through the variables $Z = \max(X, 1)$ and $\delta = I(X > 1)$ where 1 is the detection limit. We find that the distribution of Z is given by

$$F_Z(z) = P(Z \leq z) = \begin{cases} 0 & , z < 1 \\ P(X \leq z) & , z \geq 1 \end{cases} .$$

For each of the four mouse populations in the ethanol-induced anesthesia data set, and separately for each trial, we use two different graphical methods to verify whether the normal assumption is valid for ethanol-induced sleeping time. In Figures 4.1 and 4.2 we compare the fitted normal survival function with a non-parametric Kaplan-Meier estimate for left-censored data (Gomez et al (1992)). The unknown parameters μ and σ^2 in the fitted survival function are estimated via maximum likelihood and are given in Table 4.1.

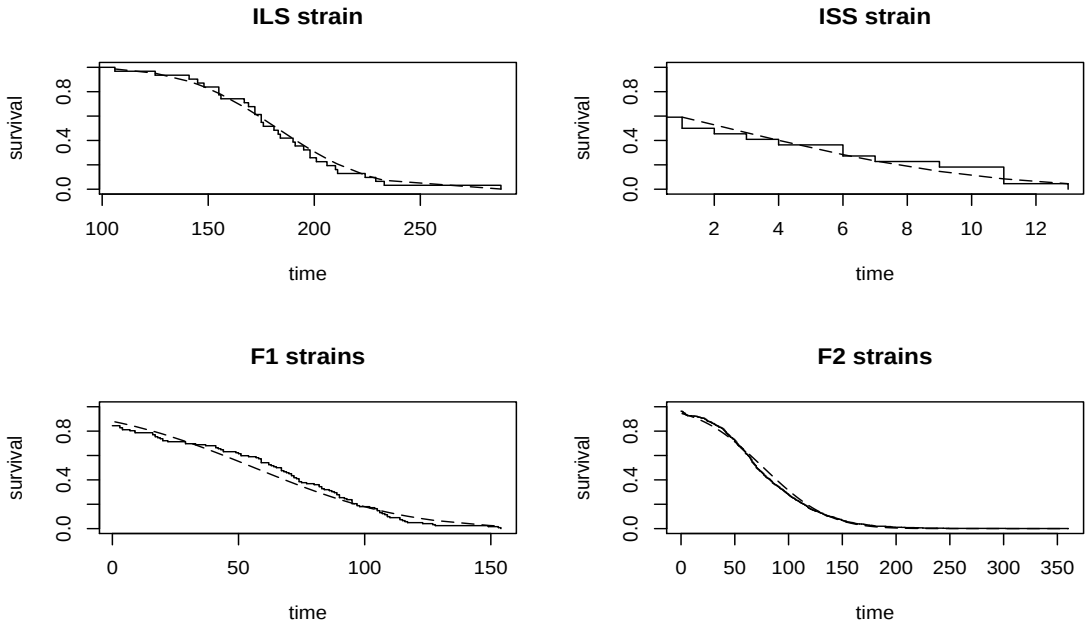


Figure 4.1. : A non-parametric Kaplan-Meier estimate (solid line) and a parametric survival function estimate under normality (dashed line) for the sleeping times in trial 1 of each mouse population.

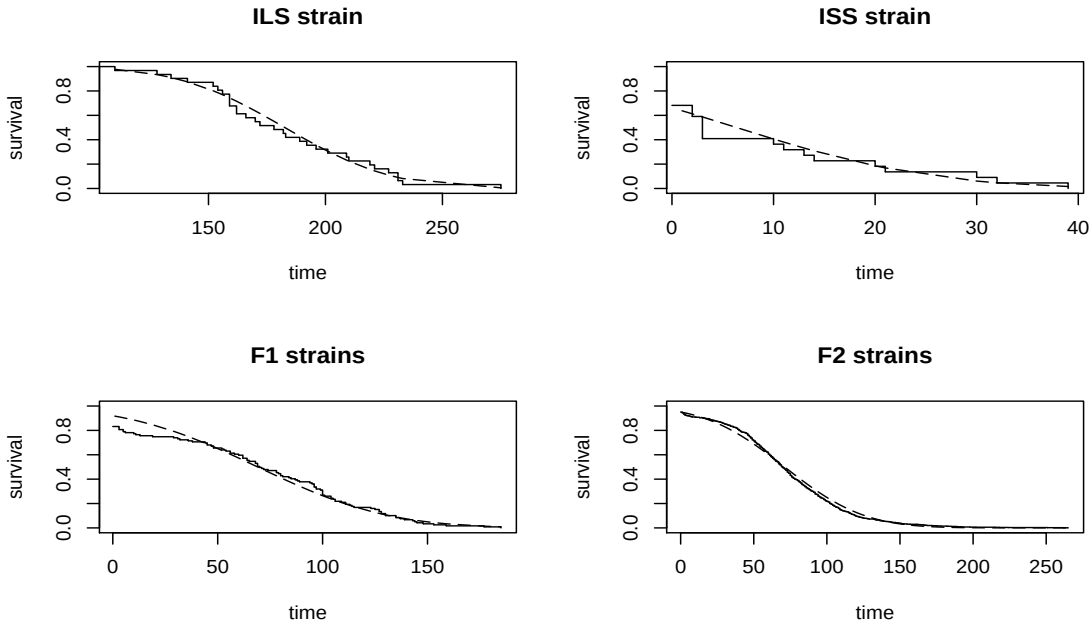


Figure 4.2. : A non-parametric Kaplan-Meier estimate (solid line) and a parametric survival function estimate under normality (dashed line) for the sleeping times in trial 2 of each mouse population.

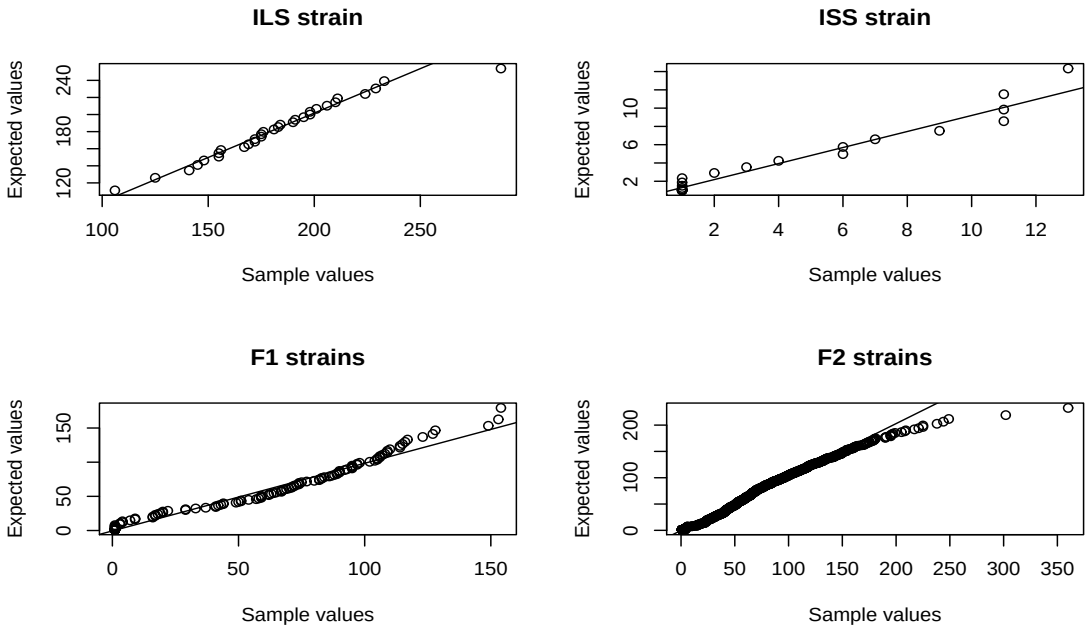


Figure 4.3. : QQ-plot for the sleeping times in trial 1 of each mouse population in the ethanol-induced anesthesia data set.

We note that the non-parametric Kaplan-Meier estimate is in each of the plots close to the fitted normal survival function. Furthermore we see in some plots that the difference between the Kaplan-Meier estimate and the fitted normal survival function is smallest for large values of sleeping time and becomes larger for smaller values of sleeping time. This result follows from the construction of the Kaplan-Meier for left-censored data which is not a proper survival function when the smallest observation is censored. A similar result is known for the Kaplan-Meier estimate for right-censored observations.

A second graphical method to verify whether the normal assumption is valid for ethanol-induced sleeping time, is to construct quantile-quantile plots (QQ-plots) for the observed data. In these plots we compare the ordered observed sample with the expected values of the order statistics under a censored normal distribution. Unlike the uncensored situation, this censored distribution cannot be transformed into a distribution which does not depend on the parameters μ and σ^2. These parameters play an important role in the expressions of the expected values of the order statistics. Therefore we estimate them by maximum likelihood, as given in Table 4.1. This estimation method is also described in Lyles et al. (2001). In Figures 4.3 and 4.4 we show the QQ-plots for the ethanol-induced sleeping time in mice. As we did in Figures 4.1 and 4.2, we give a plot for each mouse population and for each trial separately. We note in these plots that, for

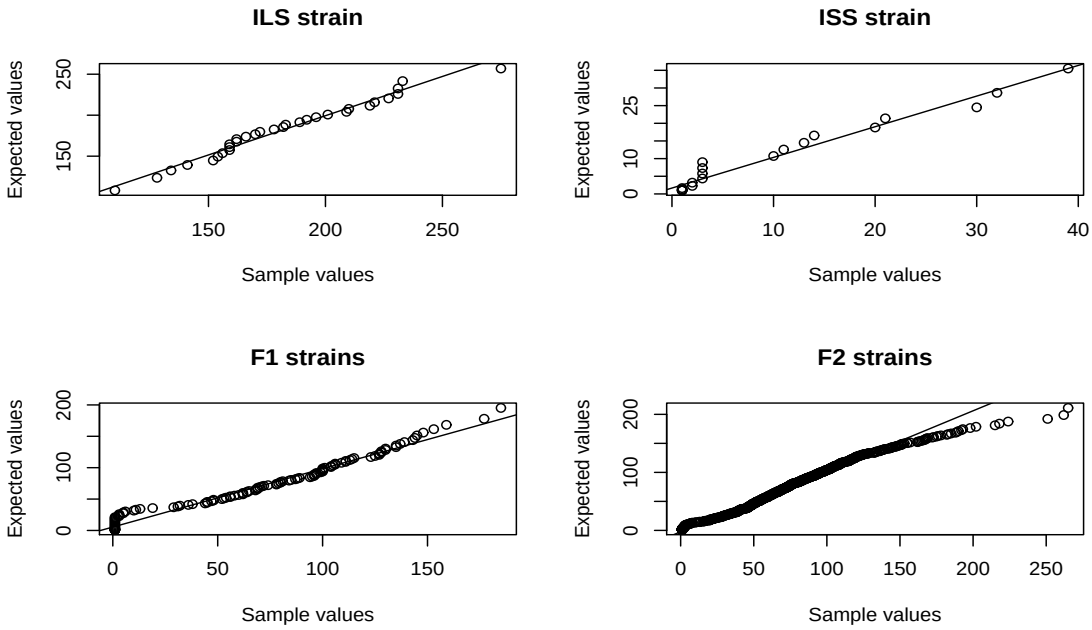


Figure 4.4. : QQ-plot for the sleeping times in trial 2 of each mouse population in the ethanol-induced anesthesia data set.

the isogenic populations ILS, ISS and F_1, almost all the observations lie on a straight line. In these populations it seems reasonable to assume a normal distribution for the ethanol-induced sleeping time. However due to the small sample size in the ILS and ISS populations, we need to be careful with this assumption. For the F_2 population we note that in both trials the majority of the observations lie on a straight line. For the largest observations we see that they systematically deviate from this line. This indicates that the underlying distribution of this population has slightly larger tails than the normal distribution. However the deviation in the upper tail is not very large, so we will work in the further analysis with the normal distribution.

In the Figures 4.1 till 4.4 above, we used the maximum likelihood estimates for the mean and variance parameters of a censored normal distribution. We give in Table 4.1, the different estimates for each of the four mice populations, separately for each trial. We note that the proportion of left-censored observations is different in each of the populations. In the ILS population, there are no censored observations while in the ISS population, 41% of measurements are censored in the first trial and 32% in the second trial. In the F_1 population the percentage of censored observations is around 16% and for the F_2 population this percentage is reduced to less than 5%. Furthermore we see that three animals in the F_1 population and two animals in the F_2 population have a missing value

Population	Trial	N	Censored (%)	ML-method $\bar{x}$	s
ILS	1	31	0.00	182.5484	34.68182
	2	31	0.00	182.6129	36.15529
ISS	1	22	40.90909	2.43935	6.22876
	2	22	31.81818	6.36617	15.26229
F_1	1	122	15.57377	56.23192	47.77404
	2	119	16.80672	69.03361	49.11090
F_2	1	1073	3.63467	76.76185	47.79585
	2	1071	4.94865	71.06635	42.93498
Total	1	1248	5.36859	75.86040	51.75211
	2	1243	6.43604	71.97027	48.37596

Table 4.1. : Descriptive statistics for sleeping time. For each population we give the maximum likelihood estimates for mean and standard deviation (ML-method).

for the sleeping time measurement in the second trial. Since this number of missing values is small in comparison with the total number of mice in these populations, we delete these five mice and work only with the complete cases. For one of these mice, we know that the alcohol injection was misplaced and therefore the measurement was unreliable.

4.2 Genetic modeling

In this section we present the results of an analysis on the ethanol-induced anesthesia data set. We model the underlying ethanol-induced sleeping time in mice by the linear mixed models which we described in Section 3. In first instance we investigate which genetic factors have a significant influence on this sleeping time. Afterwards we also consider the

Model	$-2 \times$ Loglikelihood	H_0 - model	Likelihood ratio	Df	P-value
(1) m	24748				
(2) ma	24622	(1)	126	1	< 0.0001
(3) mad	24578	(2)	44	1	< 0.0001
(4) madl	24577	(3)	1	1	0.3173
(5) madi	24577	(3)	1	1	0.3173

Table 4.2. : For different genetic models, we give the $-2 \times$ Loglikelihood- value and we compare these models by likelihood ratio tests.

Model	$-2 \times$ Loglikelihood	H_0 - model	Likelihood ratio	Df	P-value
(6)	24563	(3)	15	3	0.0018
(7)	24567	(6)	4	0,1,2,3	0.1005
(8)	24561	(7)	6	1	0.0143
(9)	24554	(8)	7	1	0.0082
(9a)	24554	(9)	0	1	1.0000
(10)	24551	(9)	3	1	0.0833
(11)	24278	(9)	276	1	< 0.0001
(12)	24245	(11)	33	3	< 0.0001

(6) mad*trial

(7) mad*trial

(8) mad*trial + albin

(9) mad*trial + albin + sex

(9a) mad*trial + albin*sex

(10) mad*trial + albin + sex + weight

(11) mad*trial + albin + sex + trialday

(12) mad*trial + albin + sex + trialday*popul

From model (7) we assume that the covariance between the measurements of ethanol-induced sleeping time within an animal of an isogenic population is zero.

Table 4.3. : For different genetic-environmental models, we give the $-2 \times$ Loglikelihood- value and we compare these models by likelihood ratio tests.

influence of environmental factors on the average sleeping time. However, as we noted before, the ethanol-induced anesthesia data set has left-censored observations. Therefore the joint likelihood function has no standard form and, more importantly, there does not exists any standard software procedure to estimate the parameters in this function. In this analysis we have implemented the joint likelihood function in the SAS procedure Proc nlmixed. This was not a straightforward operation since we have a three-level hierarchical linear mixed model with two different random effects for the F_2 population while the Proc nlmixed procedure only allows one random effect. By integrating out analytically the random effect for the repeated measurements in each animal, we find a marginal bivariate model for each of the isogenic populations and we reduce the three-level hierarchical model in the segregating F_2 population to a two-level model for which we are able to use the Proc nlmixed procedure to estimate the different parameters.

As a first step in the analysis of the ethanol-induced anesthesia data set, we concentrate on

Parameter	Model (3)	Model (6)	Model (7)
m	91.8201 (2.3763)	93.3890 (3.1957)	94.0516 (3.1066)
a	88.9631 (2.5968)	93.4234 (3.5448)	92.9939 (3.3457)
d	-33.8199 (4.4055)	-32.3791 (6.1868)	-33.4698 (5.8547)
trial 1		0.0	0.0
trial 2		-3.0825 (4.2728)	-3.8908 (4.3212)
trial*a 1		0.0	0.0
trial*a 2		-8.9935 (4.4916)	-8.6729 (4.6408)
trial*d 1		0.0	0.0
trial*d 2		-3.0221 (8.4676)	-1.5303 (8.5165)
σ^2_{ILS}	127.15 (227.67)	93.6302 (237.88)	
ε^2_{ILS}	1131.10 (287.30)	1198.20 (318.34)	1298.18 (239.19)
σ^2_{ISS}	60.0658 (47.1085)	72.7278 (47.8608)	
ε^2_{ISS}	110.87 (39.7347)	82.8001 (30.3370)	145.11 (41.9632)
$\sigma^2_{F_1}$	56.7889 (258.21)	22.8907 (228.63)	
$\varepsilon^2_{F_1}$	2766.19 (419.45)	2844.47 (368.95)	2862.44 (304.16)
$\sigma^2_{F_2}$	608.82 (61.0394)	617.40 (61.0544)	617.58 (61.0042)
$\varepsilon^2_{F_2}$	1128.10 (50.0747)	1111.30 (49.3581)	1110.98 (49.3432)
τ^2	329.37 (59.4314)	329.37 (59.4509)	329.40 (59.8585)

Table 4.4. The different parameter estimates and their standard error for the genetic model (3) and the genetic - environmental models (6) and (7).

finding the different genetic factors which have an influence on ethanol-induced sleeping time in mice. In Table 4.2 we give for different genetic models the loglikelihood value multiplied by minus two and we use likelihood ratio tests to select from this table an optimal model. We follow in this selection some guidelines from the genetic literature. The breeding value a must have a significant effect before we look at the dominance factor d and this factor also has to be significant before we continue with the epistatic interactions i and l. For each likelihood ratio test we give in Table 4.2 the degrees of freedom and the p-value. A genetic factor is considered significant when the p-value is less than 5%. We assume that for each model, the parameters in the variance structure are different for each mouse population in this data set. From Table 4.2 we note that the breeding value a and the dominance factor d are highly significant. Both the epistatic interactions i and l are non-significant. Therefore we take as optimal genetic model the model (3) with breeding value a and dominance factor d. This model will form the starting point in the development of genetic-environmental models to investigate which environmental factors have a significant influence on the sleeping time in mice.

In Table 4.3 we give the loglikelihood function, multiplied by minus two, for several genetic-environmental models. Below the table we write for each model which covariates it contains. By likelihood ratio tests we select an optimal genetic-environmental model and investigate which environmental factors have a significant influence on the ethanol-induced sleeping time in mice. We note in the table that a first significant environmental factor is trial. This factor shows that we have a different genetic structure in the different trial sessions. Furthermore we see that the variable sex and the variable albin which indicate whether a mouse in the F_2 population is an albino or not, also have a significant effect on the average sleeping time. The ethanol-induced anesthesia data set was constructed over a period of two years. This has an influence on the ethanol-induced sleeping time in the sense that we see that the day of a trial session has a significant effect. For each mouse population the influence of the trial day is different. Next to the mean structure we investigate in Table 4.3 the variance structure for each of the four mouse populations. More specifically we are interested whether the covariance between the measurements within an animal of an isogenic population is significantly different from zero. Model (7) in the table has the same mean structure as model (6) but assumes that the covariances are zero. The null distribution of the likelihood ratio test between these models is a mixture of four chisquare distributions with respectively zero, one, two and three degrees of freedom. This result is given by Verbeke and Molenberghs (2003). We note that we cannot reject the null hypothesis and that the measurements in the two trial sessions are possibly uncorrelated. Therefore we assume for the further models that the measurements of the trial sessions are uncorrelated.

To investigate in more detail which genetic and environmental factors have a significant influence on the ethanol-induced sleeping time in mice, we give in Table 4.4 and Table 4.5 the parameter estimates with a standard error for the optimal genetic model (3) and the genetic-environmental models (6), (7), (11) and (12). For each model we use Wald tests to verify which factors in the mean structure are significant. We first note that the genetic factors m, a and d are significant in each of the models. The genes, determining the ethanol-induced sleeping time in mice, have an additive and dominant effect but not an epistatic effect. This indicates that the genes at different loci probably do not interact much. In Tables 4.4 and 4.5 we see that the dominance factor d always has a negative sign such that the average sleeping time for heterozygous animals in this data set lies closer to the average sleeping time of homozygous short sleep animals than to the average sleeping time of homozygous long sleep animals. Except for model (3), we have that the additive term a is significantly different in the two trial sessions. This indicates that some genes are activated or deactivated after the first alcohol use in the first trial session and shows that the body of some mice adapted to the alcohol use. The

Parameter	Model (11)	Model (12)
m	115.56 (5.5251)	196.23 (15.9307)
a	95.5033 (3.5100)	94.3915 (53.8572)
d	-37.0798 (6.1819)	-152.82 (23.1947)
trial 1	0.0	0.0
trial 2	-3.6533 (4.3320)	-2.1448 (4.2363)
trial*a 1	0.0	0.0
trial*a 2	-8.5878 (4.7237)	-8.9577 (4.7763)
trial*d 1	0.0	0.0
trial*d 2	-1.1871 (8.5125)	-3.5145 (8.2903)
albinism	-4.7103 (2.5613)	-6.1791 (2.6128)
sex	-4.8748 (1.8881)	-4.7444 (1.8872)
trialday	-0.0347 (0.0082)	
trialday*popul ILS		-0.2216 (0.1340)
trialday*popul ISS		-0.2323 (0.1235)
trialday*popul F_1		0.0923 (0.0381)
trialday*popul F_2		-0.0702 (0.0107)
ε^2_{ILS}	1329.02 (248.19)	1293.00 (238.21)
ε^2_{ISS}	142.16 (40.8697)	135.88 (39.3352)
$\varepsilon^2_{F_1}$	2883.41 (306.94)	2772.48 (293.43)
$\sigma^2_{F_2}$	607.53 (61.1457)	611.03 (61.3227)
$\varepsilon^2_{F_2}$	1113.32 (49.7638)	1112.26 (49.6896)
τ^2	258.06 (53.6566)	218.69 (48.0044)

Table 4.5. The different parameter estimates and their standard error for the genetic - environmental models (11) and (12).

average ethanol-induced sleeping time decreases in the second trial session. In model (11) and (12) we note that, in the F_2 population, the albino mice have a significant lower average sleeping time than mice of another color. We saw this result also in Markel and Corley (1994). They believe that the gene responsible for albinism (Tyr) is also one of the genes determining the ethanol-induced sleeping time or is closely related to one of these genes. The variable sex also has a significant effect in model (11) and (12). We see that female mice have a lower sleeping time than male mice. In the same models we consider the variable trial day which indicates the day of a trial session in the study period. From Table 4.5 we have that the estimate for this variable almost always has a negative sign. This indicates that the average ethanol-induced sleeping time is decreasing over the whole testing period. A possible explanation for this phenomenon is that the

Model	Heritability h^2	Number of genes n_L	
(3)	0.3188 (0.05077)		7 (exact: 6.0072)
(6)	0.3201 (0.05095)	Trial 1	7 (exact: 6.6247)
		Trial 2	6 (exact: 5.4107)
(7)	0.3201 (0.05123)	Trial 1	7 (exact: 6.5633)
		Trial 2	6 (exact: 5.3962)
(11)	0.2608 (0.04923)	Trial 1	9 (exact: 8.8361)
		Trial 2	8 (exact: 7.3184)
(12)	0.2252 (0.04587)	Trial 1	
		Intercept	10.1855
		Trialday	0.001155 (0.02606)
		Trialday^2	3.274×10^{-8} (1.515×10^{-6})
		Trial 2	
		Intercept	8.344
		Trialday	0.001045 (0.02351)
		Trialday^2	3.274×10^{-8} (1.515×10^{-6})

Table 4.6. : The heritability with standard error and the number of effective genes for genetic model (3) and genetic-environmental models (6), (7), (11) and (12). The number of genes depends on the trial session and on the trialday (standard error between brackets).

researchers are getting more skilled in the assessment of the sleeping time and therefore can decide earlier whether a mouse is asleep or not. Sofar we have only discussed the mean structure of the different models given in Tables 4.4 and 4.5. By a likelihood ratio test we found in Table 4.3 that there is no significant difference between the likelihood values in the models (6) and (7). When we look at the estimates of the covariances between the measurements of the same animal at the different trial sessions in model (6), we note that they are smaller or not considerably larger than their standard errors. Therefore it is reasonable to assume that these parameters are zero.

Based on the parameter estimates in Tables 4.4 en 4.5 we derive for each model the heritability of the ethanol-induced sleeping time in mice. Via a delta-method we also calculate a standard error for this estimate. By combining estimates from the mean and variance structure of each model we find an estimate for the lower bound on the number of effective genes determining the ethanol-induced sleeping time. This number is an indication for the number of genes with different alleles in each of the parental strains and forms a lower bound for the true number of genes which determine the ethanol-induced sleeping time in mice. For several models this estimate depends on the

interaction between trial and breeding value or on the trial day. We show the different results in Table 4.6. In model (3), (6) and (7), we note that the heritability is about 31%, while the effective number of genes is 7. For model (6) and (7) this number changes due to the significant additive effect between the trial sessions into 6 at the second trial session. In model (11) and (12) the heritability and the number of effective genes are also influenced by the trial day. In these models the heritability is decreasing to respectively 26% and 22.5%. The number of effective genes in model (11) increases to 9 for trial session 1 and to 8 for trial session 2. In model (12) the effective number of genes is a quadratic function of the trial day, however the coefficients for the linear and quadratic terms are non-significant. The effective number of genes is 10 in the first trial session and decreases to 8 in the second trial session. The results for model (3), (6) and (7) are similar to the results found in Markel et al (1995b).

5 Conclusions

In this paper we reanalyzed the ethanol-induced data set given in Markel and Corley (1994) and in Markel et al. (1995a and 1995b). This data set consists of four different mouse populations, over three different generations. These populations were specially bred to investigate which genetic and environmental factors have an influence on the sleeping time in mice. However due to the difficulty in the assessment of the ethanol-induced sleeping time in each animal, we see that some of the observations in this data set were censored to the left by a detection limit of 1 min. Furthermore we were able to identify the litters of mice in the F_2 population.

We developed in this paper a new method to analyze the ethanol-induced anesthesia data set in which we fully took into account the complex structure of this data set. This method is an extension of the linear mixed models framework. We assumed that the underlying ethanol-induced sleeping time in mice was normally distributed and we developed in each mouse population a separate linear mixed model. The variance structure in the isogenic populations was determined by a two-level hierarchical model while we considered a three-level hierarchical model for the F_2 population. The mean structure in each of the models was related such that we were able to investigate the influence of genetic and environmental factors on the sleeping time. Based on this parameters we found estimators for the heritability and the number of effective genes for this trait.

In an exploratory data analysis we gained more insight into the data and verified the distributional assumptions of the linear mixed models. Afterwards we investigated which

genetic and environmental factors had an influence on the ethanol-induced sleeping time. For the genetic factor we found that the genetic factors breeding value a and dominance d are significant. For the environmental factors we noted first that the covariate trial session has an influence on the sleeping time of mice through its interactions with genetic factors. Some other environmental factors such as sex, albinism in the F_2 population and trial day were also significant.

From the parameter estimates in the different genetic and genetic-environmental models we calculated the heritability of ethanol-induced sleeping time and also gave a lower bound on the number of effective genes for this trait. In the majority of these models we found a heritability estimate of 32% and a number of effective genes of 7. These results correspond to those in Markel et al. (1995b). For the models depending on the factor trial day of the test sessions, the heritability slightly decreases and the number of effective genes increases to about 10.

Acknowledgement

This work was supported by the Ministry of the Flemish Community (Project BIL00/28, International Scientific and Technological Cooperation) and by the IAP research network nr P5/24 of the Belgian State (Belgian Science Policy).

References

[1] Cohen, A. C. (1959) Simplified estimators for the normal distribution when samples are singly censored or truncated, *Technometrics*, **1**, 217-237.

[2] Falconer, D. S. (1989) *Introduction to quantitative genetics*, New York: John Wiley.

[3] Gomez, G., Julià, O., and Utzet, F. (1992) Survival analysis for left censored data. In *Survival analysis: State of the art*, Ed. J. P. Klein and P. K. Goel, Nato ASI Series E: Applied Sciences, Netherlands: Kluwer, 269-288.

[4] Hughes, J. P. (1999) Mixed effects models with censored data with application to HIV RNA levels, *Biometrics*, **55**, 625-629.

[5] Lyles, R. H., Fan, D., and Chuachoowong, R. (2001) Correlation coefficient estimation involving a left censored laboratory assay variable, *Statistics in Medicine*, **20**, 2921-2933.

[6] Lynch, M. and Walsh, B. (1998) *Genetics and analysis of quantitative traits*, Sunderland: Sinauer Associates, Inc.

[7] Markel, P. D., and Corley, R. P. (1994) A multivariate analysis of repeated measures: linkage of the albinism gene (Tyr) to a QTL influencing ethanol-induced anesthesia in laboratory mice, *Psychiatric Genetics*, **4**, 205-210.

[8] Markel, P. D., DeFries, J. C. and Johnson, T. E. (1995a) Use of repeated measures in an analysis of ethanol-induced loss of righting reflex in inbred long-sleep and short-sleep mice, *Alcoholism: Clinical and Experimental Research*, **19**, 299-304.

[9] Markel, P. D., DeFries, J. C. and Johnson, T. E. (1995b) Ethanol-induced anesthesia in inbred strains of long-sleep and short-sleep mice: A genetic analysis of repeated measures using censored data, *Behavior Genetics*, **25**, 67-73.

[10] Mather, K. and Jinks, J. L. (1982) *Biometrical genetics*, Cambridge: University Press.

[11] Thiébaut, R. and Jacqmin-Gadda, H. (2004) Mixed models for longitudinal left-censored repeated measures, *Computer methods and programs in biomedicine*, **74**, 255-260.

[12] Verbeke, G. and Molenberghs, G. (2003) The use of score tests for inference on variance components, *Biometrics*, **59**, 254-262.

[13] Wright, S. (1968) *Evolution and the genetics of populations Vol. 1*, Chicago: University of Chicago Press.

[14] Zeng, Z. (1992) Correcting the bias of Wright's estimates of the number of genes affecting a quantitative character: a further improved method, *Genetics*, **131**, 987-1001.

Authors:

Prof. dr. Roel Braekers (Corresponding author) E-mail: roel.braekers@uhasselt.be
Hasselt University
Center for Statistics
Campus Diepenbeek, Agoralaan - Building D
3590 Diepenbeek
Belgium

Prof. dr. Paul D. Markel E-mail: markel@minotstateu.edu
Department of Addiction Studies, Psychology & Social Work
Minot State University
500 University Avenue West
Minot, North Dakota 58707
U.S.