

Article

Regression Models for Log-Normal Data: Comparing Different Methods for Quantifying the Association between Abdominal Adiposity and Biomarkers of Inflammation and Insulin Resistance

Sara Gustavsson ^{1,*}, Björn Fagerberg ^{2,3}, Gerd Sallsten ¹ and Eva M. Andersson ¹

¹ Occupational and Environmental Medicine, Sahlgrenska University Hospital and Academy, University Of Gothenburg, Gothenburg SE-405 30, Sweden;

E-Mails: gerd.sallsten@amm.gu.se (G.S.); eva.m.andersson@amm.gu.se (E.M.A.)

² Wallenberg Laboratory, Sahlgrenska Center for Cardiovascular and Metabolic Research, Sahlgrenska University Hospital, Gothenburg SE-413 45, Sweden;

E-Mail: bjoern.fagerberg@wlab.gu.se

³ Department of Molecular and Clinical Medicine, University of Gothenburg, Gothenburg SE-413 45, Sweden

* Author to whom correspondence should be addressed; E-Mail: sara.gustavsson@amm.gu.se; Tel.: +46-(0)31-786-3621.

Received: 31 January 2014; in revised form: 14 March 2014 / Accepted: 20 March 2014 /

Published: 27 March 2014

Abstract: We compared six methods for regression on log-normal heteroscedastic data with respect to the estimated associations with explanatory factors (bias and standard error) and the estimated expected outcome (bias and confidence interval). Method comparisons were based on results from a simulation study, and also the estimation of the association between abdominal adiposity and two biomarkers; C-Reactive Protein (CRP) (inflammation marker,) and Insulin Resistance (HOMA-IR) (marker of insulin resistance). Five of the methods provide unbiased estimates of the associations and the expected outcome; two of them provide confidence intervals with correct coverage.

Keywords: linear regression model; log-normal distribution; heteroscedasticity; biomarkers of inflammation; insulin resistance; simulation study

1. Introduction

A common objective in medical research is to identify and quantify associations. For example, this could include evaluating a biomarker or estimating personal exposure levels based on questionnaires and occupational history. In these cases regression analysis is often used. It can also be important to estimate the expected value, e.g., the expected exposure. A person's risk of developing an exposure-caused disease is related to the dose, and the dose is usually estimated by the cumulative exposure. In group-based exposure assessment, the arithmetic mean is considered superior to the geometric mean, as a dose-related variable [1,2]. The arithmetic mean is also preferred, in the form of mean exposure for individuals over time, when assessing long-term effects of exposures [3].

Many biological variables (e.g., exposure and biomarkers) have a skewed distribution with a median smaller than the mean and only positive values. It is also common with heteroscedasticity, where the variance increases with the expected value. Such data can often be described by a log-normal or quasi-log-normal distribution [4–6]. A common way to analyze a log-normal variable Y is to log-transform ($Z = \ln(Y)$) so that Z follows a normal distribution with expected value μ_Z and standard deviation σ_Z . The geometric mean of Y is then found as $\exp(\mu_Z)$, while the expected value of Y (the arithmetic mean) is found as $\mu_Y = \exp(\mu_Z + \sigma_Z^2/2)$. In cases where the expected value μ_Y depends on several predictors, regression analysis is often based on the log-transformed data, $Z = \hat{\delta}_0 + \hat{\delta}_1 x_1 + \dots + \hat{\delta}_p x_p + e$, and the expected value of Y is estimated as $\hat{\mu}_Y = \exp(\hat{\delta}_0 + \hat{\delta}_1 x_1 + \dots + \hat{\delta}_p^2/2)$. This produces effect-measures on the multiplicative scale and the interpretation is that Y is expected to increase $100(\exp(\delta_i) - 1)$ percent as x_i increases one unit, see e.g. [7].

We investigated the situation where we want an estimate of the absolute effect, thus we need the model to be linear on the original scale, $\mu_Y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$, in order to produce effect-measures on the additive scale. This is of interest e.g., in exposure modeling, when exposure time is an important factor and it is reasonable that the effect of time on exposure is linear. Effect-measures on the additive scale have also been discussed in relation to statistical vs. biologic interaction. Biologic interaction occurs when the effect of one cause depends on the presence of another cause, e.g., environmental causes and genetic predisposition, and is often defined as departure from additivity [8,9].

Different regression methods, suitable for log-normal data, were investigated and the aim was to estimate the absolute effect β_i of each predictor. Because of the heteroscedasticity, the ordinary least squares regression will produce erroneous tests and confidence intervals. One solution is to use a weighted least squares regression. Another way to handle non-normal distributions is to use a general linear model, GLM, in which the distribution of the response variable Y belongs to the natural exponential family and the expected value of Y is linked to a linear model by a link function, $g(\mu_Y) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$, see [10]. One example of a GLM that is suitable for the log-normal distribution is the gamma distribution with an identity link. Another possibility is the normal distribution and an exponential link, applied to $Z = \ln(Y)$.

We compared the different regression methods using both large scale simulations and by applying them to a cross-sectional data set with the aim to quantify the association of abdominal adiposity with inflammation and insulin resistance (two well-known associations).

2. Linear Regression with a Lognormal Response

We considered a regression model where the expected value of a continuous log-normal response variable Y is a linear function of the predictors X_1, X_2, \dots, X_p :

$$\mu_Y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p \quad (1)$$

The variance of Y depends on both the expected value of Y , μ_Y , and the variance of $Z = \ln(Y)$, σ_Z^2 ; $\text{Var}[Y|X] = \sigma_{Y|X}^2 = (\exp(\sigma_Z^2) - 1) \cdot \mu_Y^2$.

Ordinary least squares regression (here denoted LS_{lin}) can be used to obtain unbiased estimates $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p$. However, the estimates provided by LS_{lin} assume homoscedasticity, which, as previously noted, is incorrect for a log-normal variable. This incorrect variance assumption leads to incorrect statistical inferences.

In a situation with heteroscedasticity, weighted least squares regression (here denoted WLS) can be used. WLS can account for the heteroscedasticity by weighting each observation, Y_i , with the inverse of its variance, $W_i \propto \sigma_{Y_i}^{-2}$. For a log-normal distribution, the weight for Y_i is $W_i = 1/\mu_{Y_i}^2$, where LS_{lin} can provide estimates of μ_{Y_i} . Unlike LS_{lin} , WLS provides an estimate of the variance σ_Z^2 .

When the response Y is log-normally distributed, data are often log-transformed, $\ln(Y) = Z$, and a log-linear model is estimated:

$$\mu_{Z|X} = \delta_0 + \delta_1 X_1 + \dots + \delta_p X_p \quad (2)$$

where the expected value of Y is $\mu_{Y|X} = \exp(\mu_{Z|X} + \sigma_Z^2/2)$. Ordinary least squares regression on Z (here denoted LS_{exp}) provides estimates of the relative effect ($\hat{\delta}_0, \hat{\delta}_1, \dots, \hat{\delta}_p$) as well as an estimate of the variance σ_Z^2 but no estimates of the absolute effects. Thus, both (1) and (2) can be used to estimate $\mu_{Y|X}$ and σ_Z^2 . The reason for including LS_{exp} , even if the linear model in (1) is assumed, is that LS_{exp} is commonly used for log-normal data.

The log-normal distribution is often approximated by the gamma distribution, with parameters μ (expected value) and ν (scale parameter, $\text{Var}[Y] = \mu^2/\nu$). A generalized linear model (GLM) with gamma distribution and the identity link (denoted GLM_G), provides estimates $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p$ and an estimate of σ_Z^2 can be found through the transformation $\ln(1/\nu + 1) = \sigma_Z^2$.

Another GLM that can be used to estimate the absolute effects is one with a normal distribution and the link function $\exp(*)$, applied to $Z = \ln(Y)$, here denoted GLM_N , such that

$$\exp(\mu_{Z|X}) = \phi_0 + \phi_1 X_1 + \dots + \phi_p X_p \quad (3)$$

The expected value of Y is then found as $\mu_{Y|X} = \exp(\mu_{Z|X}) \cdot \exp(\sigma_Z^2/2)$.

The method GLM_N , does not, however, take into account the stochastic variation due to estimating σ_Z^2 . Therefore we also used a maximum likelihood method (ML_{LN} , see [11,12]) based on the likelihood function of the log-normal distribution:

$$f_Y(y) = \frac{1}{y} \frac{1}{\sqrt{2\pi}\sigma_Z} \exp\left[-\frac{(\ln(y) - \mu_Z)^2}{2\sigma_Z^2}\right] \quad (4)$$

where $\mu_z = \ln(\beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip}) - \sigma_z^2 / 2$. The estimates $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p$ and $\hat{\sigma}_z^2$ are found using iterations, for example the Newton-Raphson iteration used here [13].

2.1. Confidence Intervals

For LS_{lin} , WLS , GLM_G and ML_{LN} , a 95% confidence interval for $\mu_{Y|X}$ is estimated as $\mu_{Y|X} \pm z_{\alpha/2} \sqrt{\text{var}(\hat{\mu}_{Y|X})}$, where the sample-specific variance is estimated as:

$$\text{var}(\hat{\mu}_{Y|X}) = \sum_{i=0}^p x_i^2 \cdot \text{var}(\hat{\beta}_i) + 2x_0 x_1 \cdot \text{cov}(\hat{\beta}_0, \hat{\beta}_1) + \dots + 2x_{p-1} x_p \cdot \text{cov}(\hat{\beta}_{p-1}, \hat{\beta}_1) \tag{5}$$

where $x_0 = 1$, $\text{var}(\hat{\beta}_i)$ and $\text{cov}(\hat{\beta}_i, \hat{\beta}_j)$ are the sample-specific estimates of the variance and the covariance (the sample-specific standard error is $se(\hat{\beta}_i) = \sqrt{\text{var}(\hat{\beta}_i)}$).

For GLM_N , a confidence interval is estimated as $(\exp(\mu_{Y|X}) \pm z_{\alpha/2} \sqrt{\text{var}(\exp(\hat{\mu}_{Y|X}))}) \cdot \exp(\hat{\sigma}_z^2 / 2)$, where the sample-specific variance of the linear estimator is estimated as:

$$\text{var}(\exp(\hat{\mu}_{Z|X})) = \sum_{i=0}^p x_i^2 \cdot \text{var}(\hat{\phi}_i) + 2x_0 x_1 \cdot \text{cov}(\hat{\phi}_0, \hat{\phi}_1) + \dots + 2x_{p-1} x_p \cdot \text{cov}(\hat{\phi}_{p-1}, \hat{\phi}_1) \tag{6}$$

For LS_{exp} , a confidence interval for $\mu_{Y|X}$ is estimated as $\exp\left(\hat{\mu}_{Z|X} + \frac{\hat{\sigma}_Z^2}{2} \pm t_{\alpha/2, (n-p-1)} \sqrt{\text{var}(\hat{\mu}_{Z|X})^2 + \frac{(\hat{\sigma}_Z^2)^2}{2(n-p-1)}}\right)$, using the modified Cox method [14].

The sample-specific variance is estimated as:

$$\text{var}(\hat{\mu}_{Z|X}) = \sum_{i=0}^p x_i^2 \cdot \text{var}(\hat{\delta}_i) + 2x_0 x_1 \cdot \text{cov}(\hat{\delta}_0, \hat{\delta}_1) + \dots + 2x_{p-1} x_p \cdot \text{cov}(\hat{\delta}_{p-1}, \hat{\delta}_1) \tag{7}$$

where $x_0 = 1$, $\text{var}(\hat{\delta}_i)$ and $\text{cov}(\hat{\delta}_i, \hat{\delta}_j)$ are the sample-specific estimates of the variance and the covariance.

2.2. Simulation Model

In a simulation study we compared the large-sample properties of the methods for estimating the expected value of Y and the effect of each predictor, when data follow a log-normal distribution. To obtain a realistic scenario, a simulation model was estimated from a real-life data set on personal exposure to $PM_{2.5}$ -particles in Sweden. These data are described in [15]. $PM_{2.5}$ is the mass (microgram/ m^3) of particles smaller than 2.5 micrometers, which implies that they are small enough to bypass the respiratory defenses and enter into the lungs. Increased levels of $PM_{2.5}$ have been associated with increased mortality from cardiovascular disease and lung cancer [16,17]. Several sources contribute to the personal exposure to $PM_{2.5}$, two of them are tobacco smoke and traffic exhaust [18].

The expected outcome, personal exposure to $PM_{2.5}$ -particles ($\mu g/m^3$), was assumed to be a linear function of the number of cigarettes per day, Smoke, and residential outdoor concentration of $PM_{2.5}$ ($\mu g/m^3$), ConcOut:

$$E[Y] = \mu_Y = 1.564 + 0.122 \cdot \text{Smoke} + 0.075 \cdot \text{ConcOut} \quad (8)$$

Observations were then simulated according to the model $Z = \ln(\mu_Y) - 0.383^2/2 + \varepsilon$, where $\varepsilon \sim N(0, \sigma_Z = 0.383)$. In order to facilitate interpretation and comparison without the introduction of unnecessary variation, balanced data were used in the simulations, with the following values of the explanatory variables: $\text{ConcOut} = \{2, 8, 14\}$, $\text{Smoke} = \{0, 7, 14\}$. Thus we estimated the expected $\text{PM}_{2.5}$ exposure for 9 combinations of outdoor concentration and cigarettes smoked. Simulations with 10,000 replicates were used to evaluate the potential bias in the estimates of β_0 , β_1 and β_2 , the sample-specific standard error $se(\hat{\beta}) = \sqrt{\text{var}(\hat{\beta})}$ as well as the true standard deviation $\text{SD}[\hat{\beta}] = \sqrt{\text{Var}[\hat{\beta}]}$ and also the properties of confidence intervals for μ_Y .

2.3. The DIWA Data Set

The DIWA dataset is a population-based cohort of 64-year-old women from the city Gothenburg in Sweden and has previously been described in detail in [19]. Of the 2,595 women who was screened 9.5% had diabetes mellitus (DM) [20], and of these 230 participated in the study, together with similar sized, randomly-selected groups of women with impaired glucose tolerance (IGT, $n = 209$) and normal glucose tolerance (NGT, $n = 190$). The World Health Organization criteria for capillary glucose cut-off values were used to define diabetes and impaired glucose tolerance [21]. Insulin resistance was also assessed, as well as a large number of biomarkers including high sensitivity C-reactive protein (hS-CRP). The examination also included a questionnaire regarding medical history and lifestyle factors, including smoking habits (never smoker, past smoker and smoker) and recreational physical activity (<2 h/week and ≥ 2 h/week). Body weight and waist circumference were also measured.

CRP is an acute-phase protein found in blood serum and its levels increase during an inflammatory process. CRP is mainly used as an inflammatory marker in clinical practice and should, for a healthy person, be less than 5 mg/L. Diabetes, smoking, obesity and insulin resistance are all been associated with small increases in CRP-levels as assessed by high sensitivity methods [22–25].

Insulin resistance is a condition where the body has a reduced ability to respond to the insulin hormone which can cause blood glucose to rise above normal levels. Insulin resistance can lead to type 2 diabetes and cardiovascular disease. Even if insulin resistance is most common among persons with diabetes mellitus of type 2 or impaired glucose tolerance, it is also present in about 25% of non-obese persons with normal glucose tolerance, [26]. Obesity, and in particular abdominal obesity, is associated with increased insulin resistance [27,28]. Other factors are smoking and low physical activity [29,30]. In our study, insulin resistance was measured using the homeostasis model assessment of insulin resistance (HOMA-IR), which is a mathematical formula for quantifying insulin resistance [31]; HOMA-IR is the product of fasting serum glucose and fasting serum insulin (fasting serum glucose (mmol/L)·fasting serum insulin/22.5). A cut-off value around 2.5 is often used as an upper limit for normal HOMA-IR [32–35].

3. Results

3.1. Bias and Standard Deviation of the Regression Coefficients (Simulation Study)

In the simulation study, balanced data sets were computer-generated using the model in Section 2.2, with two explanatory variables (Smoke and ConcOut) each with three levels. To obtain a balanced sample with at least 100 observations, the sample size $n = 108$ was used. For each sample, coefficients of the regression model were estimated, along with the expected outcome (personal exposure) and its confidence interval.

Table 1. Estimates of the regression coefficients; expected value of the estimate, $E[*]$, true standard deviation of the estimated coefficient, $SD[*]$, and expected sample specific standard error, $E[se(*)]$. The true coefficient values are $\beta_0 = 1.564$, $\beta_1 = 0.122$, $\beta_2 = 0.075$, $\sigma_Z = 0.383$. Results of the simulation study for sample size $n = 108$ ($r = 10,000$ replicates).

	LS _{lin}	WLS	ML _{LN}	GLM _G	GLM _N ¹	LS _{exp} ²
Intercept						
$E[*]$	1.566	1.560	1.563	1.565	1.567	0.487
$SD[*]$	0.226	0.190	0.183	0.187	0.180	0.083
$E[se(*)]$	0.269	0.187	0.180	0.178	0.179	0.084
Parameter for X ₁						
$E[*]$	0.121	0.122	0.122	0.122	0.121	0.042
$SD[*]$	0.021	0.019	0.019	0.020	0.019	0.006
$E[se(*)]$	0.021	0.019	0.018	0.018	0.018	0.006
Parameter for X ₂						
$E[*]$	0.075	0.075	0.075	0.075	0.075	0.027
$SD[*]$	0.024	0.021	0.021	0.021	0.02	0.008
$E[se(*)]$	0.024	0.021	0.020	0.020	0.02	0.008
$E[\hat{\sigma}_{Y X}]$	1.229					
$SD[\hat{\sigma}_{Y X}]$	0.143					
Scale parameter				7.330	0.377	
SD[scale parameter]				1.015	0.026	
$E[\hat{\sigma}_Z]$		0.379	0.376	0.358 ³	0.377	0.384
$SD[\hat{\sigma}_Z]$		0.031	0.026	-	0.026	0.026

¹ After transformation of the coefficients in eq (3): $\hat{\phi}_j \cdot \exp(\hat{\sigma}_Z^2 / 2)$ and $se(\hat{\phi}_j) \cdot \exp(\hat{\sigma}_Z^2 / 2)$;

² Coefficients $\hat{\delta}_i$ estimated under assumption of a log-linear model;

³ After transformation: $\sigma_Z = \sqrt{\ln(1/\nu + 1)}$.

All methods except LS_{exp} provided unbiased estimates of the regression coefficients. Among the absolute-effects methods, GLM_N tended to have the best precision (smallest SD). The sample-specific standard errors, se , were close to the true standard deviations, SD . All methods except LS_{lin} provided reasonable estimates of σ_Z , although the transformed scale parameter from GLM_G was too small (Table 1).

All methods except LS_{exp} provided an unbiased estimate of the expected value. The interval length was similar between WLS, ML_{LN}, GLM_G and GLM_N, but tended to be smaller for the two GLM methods (Table 2).

Table 2. Estimated expected value and expected length of 95% confidence interval for $\hat{\mu}_Y$, for a sample of $n = 108$ observations (results from simulation with $r = 10,000$ replicates).

Expected value	E[$\hat{\mu}_Y$]						E[length]					
	μ_Y	LS _{lin}	WLS	ML _{LN}	GLM _G	GLM _N	LS _{exp}	LS _{lin}	WLS	ML _{LN}	GLM _G	GLM _N
1.714	1.72	1.71	1.71	1.72	1.72	1.85	0.927	0.631	0.609	0.594	0.6	0.544
2.164	2.17	2.16	2.16	2.17	2.17	2.17	0.733	0.533	0.518	0.501	0.507	0.506
2.614	2.62	2.61	2.61	2.62	2.62	2.55	0.927	0.825	0.797	0.774	0.783	0.749
2.568	2.57	2.57	2.57	2.57	2.57	2.49	0.733	0.605	0.588	0.567	0.574	0.58
3.018	3.02	3.02	3.02	3.02	3.02	2.91	0.464	0.467	0.462	0.437	0.443	0.439
3.468	3.47	3.47	3.47	3.47	3.47	3.42	0.733	0.763	0.743	0.715	0.723	0.798
3.422	3.42	3.42	3.42	3.42	3.42	3.34	0.927	0.950	0.920	0.89	0.9	0.982
3.872	3.87	3.87	3.87	3.87	3.87	3.92	0.733	0.850	0.827	0.796	0.804	0.914
4.322	4.32	4.32	4.32	4.32	4.32	4.60	0.927	1.026	0.997	0.962	0.972	1.351

LS_{lin} had the largest standard deviation, especially for small and large values of μ_Y . Among the methods that provided an unbiased estimate of μ_Y , GLM_N had the smallest standard deviation. For all methods except LS_{lin}, the sample-specific standard error tended to be an underestimation ($SD[\hat{\mu}_Y] > E[se(\hat{\mu}_Y)]$), Table 3.

Table 3. True standard deviation and sample-specific standard error for the $\hat{\mu}_Y$ -values; $SD[\hat{\mu}_Y] = \sqrt{\text{Var}[\hat{\mu}_Y]}$ and $se(\hat{\mu}_Y) = \sqrt{\text{var}(\hat{\mu}_Y)}$. Results from simulation with $n = 108$ observations, $r = 10,000$ replicates.

Expected value	SD[$\hat{\mu}_Y$]						E[$se(\hat{\mu}_Y)$]					
	μ_Y	LS _{lin}	WLS	ML _{LN}	GLM _G	GLM _N	LS _{exp}	LS _{lin}	WLS	ML _{LN}	GLM _G	GLM _N
1.714	0.191	0.161	0.156	0.159	0.154	0.136	0.238	0.159	0.154	0.152	0.154	-
2.164	0.145	0.135	0.132	0.135	0.132	0.128	0.188	0.135	0.131	0.128	0.13	-
2.614	0.220	0.209	0.202	0.211	0.205	0.190	0.238	0.208	0.201	0.198	0.201	-
2.568	0.167	0.153	0.150	0.154	0.151	0.147	0.188	0.153	0.148	0.145	0.147	-
3.018	0.121	0.118	0.118	0.120	0.120	0.112	0.119	0.118	0.117	0.112	0.113	-
3.468	0.210	0.195	0.190	0.196	0.192	0.204	0.188	0.192	0.187	0.183	0.185	-
3.422	0.251	0.241	0.234	0.244	0.238	0.251	0.238	0.240	0.232	0.228	0.231	-
3.872	0.228	0.217	0.212	0.219	0.215	0.235	0.188	0.214	0.209	0.204	0.206	-
4.322	0.290	0.263	0.256	0.264	0.258	0.345	0.238	0.259	0.251	0.246	0.249	-

All methods except LS_{lin} and LS_{exp} provided coverage close to the nominal, but both GLM_G and GLM_N tended to give too low coverage, whereas ML_{LN} was slightly better. Using LS_{lin} resulted in too high coverage for low values of μ_Y , and too low coverage for large values. LS_{exp} provided too low coverage both for low and high values (Table 4).

Table 4. Actual coverage of the 95% confidence interval for μ_Y based on the sample-specific standard error (results from simulation with $n = 108$ observations and $r = 10,000$ replicates).

Expected value	Coverage ¹					
	μ_Y	LS _{lin}	WLS	ML _{LN}	GLM _G	GLM _N
1.714	0.98	0.94	0.95	0.93	0.94	0.83
2.164	0.99	0.95	0.95	0.93	0.94	0.95
2.614	0.96	0.95	0.95	0.93	0.94	0.93
2.568	0.97	0.95	0.95	0.93	0.94	0.90
3.018	0.94	0.95	0.95	0.93	0.93	0.83
3.468	0.92	0.95	0.95	0.93	0.94	0.94
3.422	0.93	0.95	0.95	0.92	0.93	0.93
3.872	0.89	0.94	0.95	0.93	0.94	0.95
4.322	0.89	0.94	0.94	0.93	0.94	0.87

¹ Proportion of replicates where 95% confidence interval covers true expected value μ_Y

3.2. Application of the Regression Methods to the DIWA Dataset

The DIWA dataset consists of data from approximately 600 women for which a large amount of data, related to diabetes and obesity, were collected. Descriptive statistics for CRP, waist circumference and HOMA-IR are presented in Table 5, separate for each glucose tolerance group.

Table 5. Descriptive statistics for C-reactive protein (CRP), insulin resistance (HOMA-IR) and waist circumference.

Group	n	CRP			HOMA-IR			Waist circumference (cm)		
		Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
NGT ¹	185	2.107	1.184	2.550	1.141	0.960	0.647	88.295	88.50	8.948
IGT ¹	195	2.583	1.380	3.783	1.816	1.430	1.268	92.677	92.50	11.882
DM ¹	218	4.468	1.856	10.255	4.677	2.835	5.842	98.083	98.00	12.631

¹ Results for women with normal glucose tolerance (NGT), impaired glucose tolerance (IGT) and diabetes mellitus (DM).

3.2.1. Regression Models for C-Reactive Protein (CRP) and Insulin Resistance (HOMA-IR)

For CRP, the start model in the multivariable regression analysis included smoking, physical activity, waist circumference (WC), insulin resistance (HOMA-IR) and glucose tolerance (GT), where GT was classified into three categories: normal glucose tolerance, impaired glucose tolerance and diabetes mellitus. We used a model that allowed for different associations for the GT groups, by including the interaction terms WC·DM and WC·IGT. The final model, based on backward elimination using ML_{LN}, contained WC and HOMA-IR, but no interaction term, thereby implying that the association with WC could be similar for the three GT groups (Figure 1).

For HOMA-IR, the start model in the multivariable regression analysis included WC, physical activity and smoking, and we allowed for possible different association with WC for the different glucose groups by including the interaction between waist circumference and glucose tolerance.

The final model, based on backward elimination using ML_{LN} , contained WC and the interaction between WC-GT, thus allowing different WC parameters for each GT group (Figure 2).

Figure 1. The parameter estimates and 95% confidence intervals for the different regression methods, when estimating CRP as a function of waist circumference (WC) and HOMA-IR, using $n = 598$ observations.

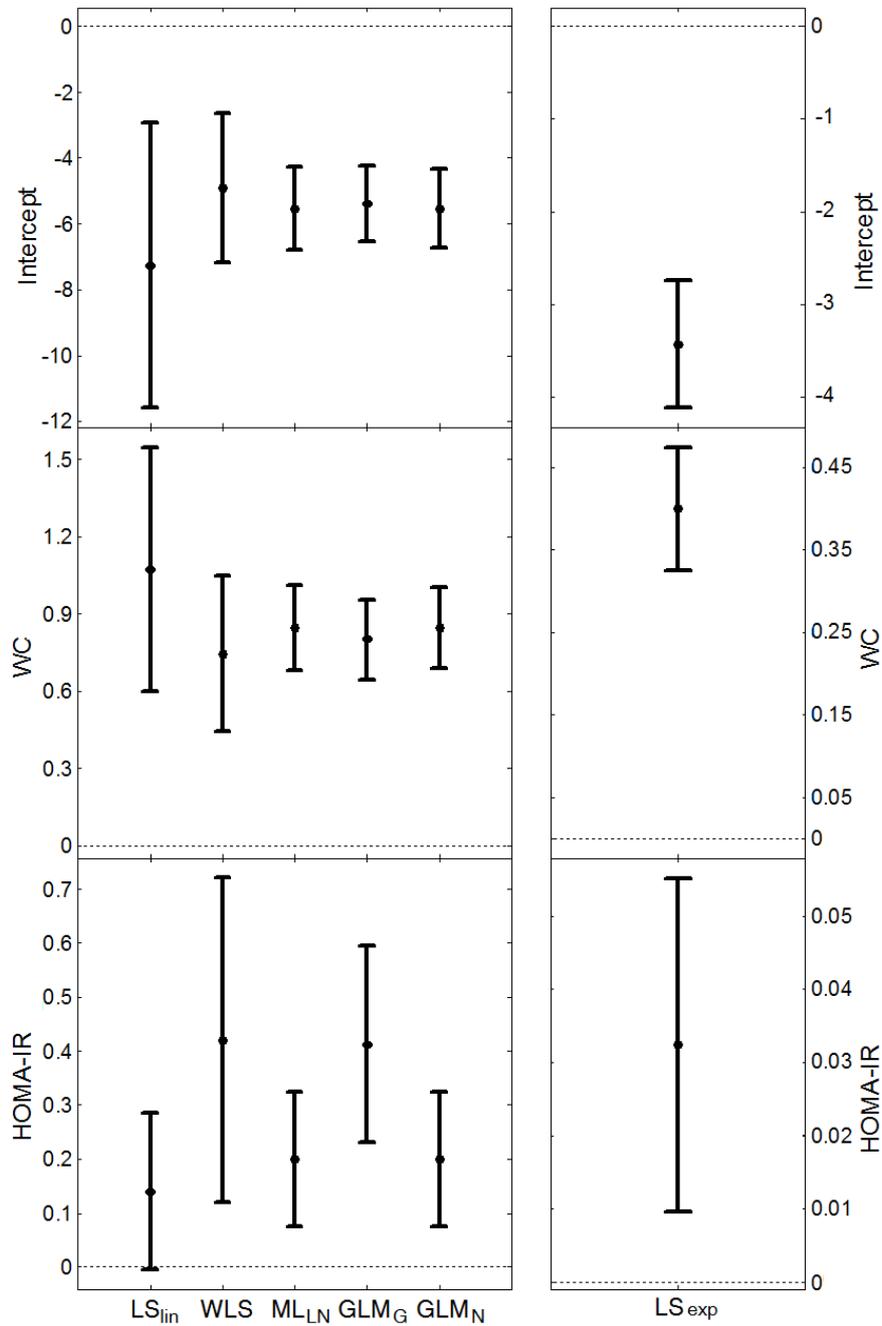
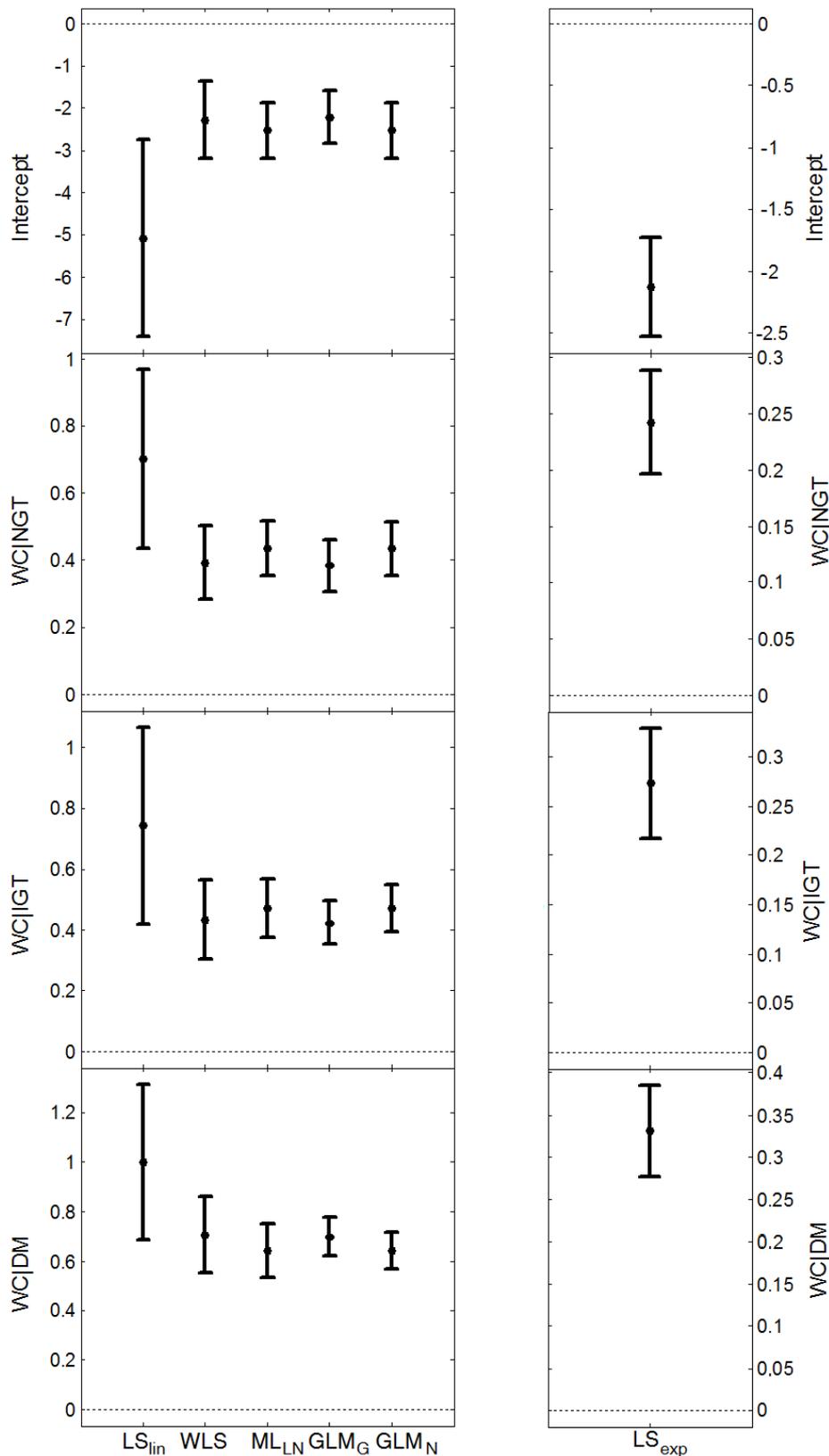


Figure 2. The parameter estimates and 95% confidence intervals for the different regression methods, when estimating HOMA-IR as a function of waist circumference (WC) and the interaction between WC and glucose tolerance group (normal glucose tolerance, impaired glucose tolerance and diabetes mellitus), using $n = 598$ observations.



The estimated standard deviation, $\hat{\sigma}_Z$, and the average length of the confidence intervals for μ_Y , (estimated from the models presented in Figures 1 and 2), are given in Table 6. ML_{LN} , GLM_N and LS_{exp} gave similar estimates of σ_Z (this parameter cannot be estimated by LS_{lin}). WLS provided the largest estimate whereas GLM_G gave the smallest. ML_{LN} and GLM_G had similar confidence intervals for the expected value, μ_Y , GLM_N had the shortest intervals, whereas LS_{lin} had the longest intervals.

Table 6. The σ_Z -estimates and mean length of 95% confidence intervals for μ_Y , for CRP and HOMA-IR, $n = 598$.

Method	CRP		HOMA-IR	
	$\hat{\sigma}_Z$	Length CI (mean, SD)	$\hat{\sigma}_Z$	Length CI (mean, SD)
LS_{lin}	-	1.61 (0.89)	-	1.10 (0.19)
WLS	1.22	1.51 (2.07)	0.73	0.64 (0.35)
ML_{LN}	1.04	0.82 (0.86)	0.61	0.43 (0.19)
GLM_G	0.71 (0.974 ¹)	0.85 (1.26)	0.33 (2.52 ¹)	0.47 (0.26)
GLM_N	1.04	0.43 (0.23)	0.61	0.23 (0.06)
LS_{exp}	1.04	1.19 (5.40)	0.60	0.50 (0.45)

¹ Estimated scale parameter

3.2.2. Quantification of Factors Associated with CRP and HOMA-OR (Method Comparison)

All of the methods demonstrated that WC was a significant predictor for CRP. According to the absolute-effects methods (LS_{lin} , WLS, GLM_G , GLM_N and ML_{LN}), the CRP was expected to increase about 1 mg/L (between 0.74 and 1.07 mg/L) for every 10 cm in WC and, according to the relative-effects method (LS_{exp}), the expected increase was 49% for every 10 cm in WC ($exp(0.40) - 1 = 0.49$), Figure 1. All methods showed a positive association between HOMA-IR and CRP. The expected increase in CRP was between 0.12 and 0.42 mg/L for every unit increase of HOMA-IR in the absolute-effects methods and 3% per unit of HOMA-IR for the relative-effects method. The association with HOMA-IR was not significant for LS_{lin} and very high for GLM_G and WLS (0.41 and 0.42, respectively). The point estimates from all methods had the same sign and for the absolute-effects methods the confidence intervals for β_{WC} overlapped, as did the intervals for $\beta_{HOMA-IR}$, Figure 1.

All methods found a positive association between HOMA-IR and WC in all glucose tolerance groups, Figure 2. Further, the results showed that women with DM had a significantly stronger association with WC than women with NGT, and this was significant for all methods. The results also indicated a stronger association with WC for women with IGT, compared to women with NGT; the interaction term for WC.IGT was significant for all absolute-effects methods except LS_{lin} . Among the absolute-effects methods, HOMA-IR was expected to increase 0.64–1.00 per 10 cm WC for women with DM, 0.42–0.74 for women with IGT and 0.39–0.70 for women with NGT. The relative-effects method showed an expected increase in HOMA-IR of 39% per 10 cm for women with DM, 31% for women with IGT and 27% for women with NGT.

4. Discussion

Several methods for estimating a linear regression on log-normal data were compared. Much research has investigated making inferences, including confidence interval, of the expected value of a log-normal distribution, e.g. [36–40]. Here we considered the situation where the systematic part of the model for the outcome Y should be additive on the original scale ($\mu_{Y|X} = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$). Had we made the assumption that the systematic part was multiplicative, the regression coefficients could have been estimated either with a GLM using gamma distribution and the log link, or by a GLM using a normal distribution and identity link for $Z = \ln(Y)$, which give similar results [41,42]. But we wanted a model for estimating the absolute effect of each explanatory factor. In exposure assessment, we often want to assess the personal exposure to e.g., a specific compound in the air, by using a model that includes the important exposure determinants. Here the quantity is an important factor (e.g., time spent in different micro-environments, number of cigarettes smoked) and it is reasonable that the effect is linear. A linear model can also be used to estimate biologic interaction, discussed in Section 4.3 below.

Six methods were compared; four of them directly modeled the expected value of Y as a linear function of the explanatory variables, $\mu_{Y|X} = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$ one method transformed the estimated coefficients, $\mu_{Y|X} = (\phi_0 + \phi_1 X_1 + \dots + \phi_p X_p) \cdot \exp(\sigma_z^2 / 2)$ and finally the common method based on log-transformation was included for comparison, $\mu_{Z|X} = \delta_0 + \delta_1 X_1 + \dots + \delta_p X_p$. Evaluation was made both using simulations and by applying the methods to a large data set to estimate well-known associations of abdominal adiposity (waist circumference, WC) on inflammation (measured using C-reactive protein, CRP) and insulin resistance (measured using HOMA-IR), respectively.

4.1. Method Comparison

In a simulation study we evaluated the regression methods in a situation where the expected outcome is a linear function of two explanatory variables. All methods except LS_{exp} provided unbiased estimates of the regression coefficients and the expected outcome, but the sample-specific standard error, $se(\hat{\beta}_i)$, tended to be too small, thus overestimating the power. For LS_{lin} , the assumption of a constant variance for Y resulted in confidence intervals for μ_Y with unnecessary high coverage for small μ_Y -values and too low coverage at large μ_Y -values. LS_{exp} does estimate the relative effect rather than the absolute and as a result the estimated expected values were biased and the coverage of the confidence intervals was erroneous. The confidence intervals from the GLM_G method had too low coverage, as a result of the underestimation of the variance σ_z^2 . This is contrary to the situation with a multiplicative model, where the gamma distribution often provide reasonable estimates when applied to a log-normal variable [41,42]. ML_{LN} , WLS and GLM_N provided approximately correct coverage, although GLM_N had a tendency to underestimate, as a result of using the estimate $\hat{\sigma}_z^2$, thus not including the stochastic variation of $\hat{\sigma}_z^2$ in the interval estimation. An approximate confidence interval taking into account its stochastic variation could be derived using Taylor expansion, see e.g. [43].

The methods were applied to two approximately log-normal response variables, CRP and HOMA-IR (almost 600 observations). The model for CRP contained WC and HOMA-IR, and the model for HOMA-IR contained WC and the interaction between WC and glucose tolerance groups

(normal glucose tolerance [NGT], impaired glucose tolerance [IGT] and diabetes mellitus [DM]). When comparing confidence intervals for β and for μ_Y , ML_{LN} and GLM_N consistently had narrower confidence intervals than WLS (and LS_{lin}). From the simulation we saw that WLS tends to overestimate the variance. Because of underestimation of σ_Z^2 , GLM_G had narrower intervals than ML_{LN} and GLM_N for μ_Y , but from the simulation we know that the coverage will be too low. Thus ML_{LN} will have a higher power and for lognormal data the probability of detecting a true explanatory variable is higher. The smaller interval lengths of ML_{LN} corroborate the results of a previous simulation study [11].

4.2. Factors Associated with CRP and HOMA-IR, Respectively

Using all methods, the analysis demonstrated a significant positive association between CRP and WC. Associations between CRP and several measures of obesity and abdominal adiposity have been shown in a number of studies [44–47], and some studies indicate that abdominal adiposity has a stronger association with inflammation than total adiposity [48–50]. For CRP we could not find any significant interaction between glucose tolerance group and waist circumference, thus our results did not indicate that the association between obesity and the inflammation marker depends on the degree of glucose tolerance. Many studies have been based on only one or two of the GT groups, [24,51–53]. Our study showed an expected increase in CRP of between 0.74 and 1.07 mg/L per 10 cm increase in WC for the absolute-effects methods and 49% per 10 cm for the relative-effects method. All methods, with the exception of LS_{lin} , showed a significant positive association between CRP and HOMA-IR. The lack of significant association using LS_{lin} can probably be explained by the estimates of the variance. In the LS_{lin} method the heteroscedasticity is not taken into account.

In the analysis of HOMA-IR, all methods identified WC as a significant predictor for HOMA-IR. There was also a significant interaction between glucose tolerance group and waist circumference, thus the absolute-effects models showed a departure from additivity. These results cannot be interpreted causally, but the interaction indicates that obesity might affect insulin resistance more for women who have diabetes mellitus compared to those with normal glucose tolerance. All models methods found a significantly stronger WC-association for women with DM compared to women with NGT, and all methods (apart from LS_{lin}) also had a significantly stronger WC-association for women with IGT compared to NGT. From the simulation we know that LS_{lin} has larger standard errors than the other methods and thus lower power. The relative-effects method LS_{exp} also showed a significant interaction between glucose tolerance group and waist circumference, *i.e.*, departure from multiplicativity.

Even if HOMA-IR typically has a skewed non-normal distribution, regression analyses have been performed using both untransformed and log-transformed HOMA-IR values, see [54,55] shows an expected increase in HOMA-IR with 3.5 units per 10 cm WC, using LS_{lin} on persons with DM, to be compared with 0.64–1.00 units in our study. The difference in association might be explained by the fact that the previous study included both men and women of different ages [56] uses the method here denoted LS_{exp} and finds a positive association; about 22% per 10 cm WC, while we found the association to be stronger; 27%–39%.

4.3. Model Choice

The choice between an additive or multiplicative model affects the interpretation of the estimated coefficients. The aim of a regression analysis might be simply to test whether there is a significant association between an outcome and a potential explanatory variable. Another aim can be to quantify a specific association (e.g., the absolute or relative effect), or assess the biologic interaction. If the study is purely exploratory, using epidemiological data, residual analysis can be used to decide which model that fits the data best. The model choice might be based on previous knowledge, e.g., about the biological process, from experimental studies.

In risk-modeling, a log-linear model is often used, $\varphi(Z, \beta) = \exp(\alpha_0 + \alpha_1 X_1 + \dots + \alpha_k X_k + \beta Z)$, where φ can be the odds ratio or rate ratio function, X_1 - X_k are covariates and Z is the exposure variable of interest. In this model the ratio has an exponential dependence on Z ; $\exp(\beta Z)$. However, linear models have also been discussed, see [57], for example in radiation epidemiology, where the linear relative rate model $\varphi(Z, \beta) = \exp(\alpha_0 + \alpha_1 X_1 + \dots + \alpha_k X_k)(1 - \beta Z)$ allows the rate ratio to increase linearly with the dose Z [58].

Not only the main effects but also potential interactions can be of interest. Interaction in a statistical sense is scale dependent, e.g., an absence of interaction in absolute-scale will lead to interaction in log-scale. An interaction in a linear absolute-effects model is additive, while an interaction in a log-linear relative-effects model is multiplicative. In epidemiology, an additive interaction (effect-modification on the absolute scale) is often considered more important when assessing public health impact, and seems to correspond more to biologically based notions of interaction [9,59,60]. There is a need for regression methods that can assess biologic interaction, as discussed in several articles. In logistic regression it is implicit that we have a multiplicative statistical relation and if an additive biological model holds, the logistic analysis would require three parameters to summarize the joint effects of only two variables, [61]. Additive interactions are given directly in a linear model, however a logistic regression model can be defined in such a way that additive interactions (e.g., biologic interaction) can be assessed [62].

4.4. Strengths and Weaknesses

Five regression methods for estimating associations on the absolute scale of the explanatory variables were compared, with regard to bias and standard deviation for the estimated coefficients and also with regard to the estimated expected outcome and its confidence interval. In addition, the standard method for log-normal data (log-transformation) was evaluated. The comparison of the methods was made both in a simulation study and using two examples. The absolute-effects methods provide similar results for the association with the predictors for CRP and HOMA-IR, respectively. The results from the examples are consistent with those from the simulations.

The aim of this study was not to provide a complete statistical model of which factors that are associated with CRP and HOMA-IR, but to compare the statistical methods. The number of factors in the regression models was therefore kept small; the simulation model only included two explanatory variables and in the models for CRP and HOMA-IR, only those variables that were significant after backward elimination using ML_{LN} were included. Thus, all factors were significant for ML_{LN} (and also

for GLM_N). This could be seen as an advantage for these methods, compared to for example a situation in which LS_{exp} had been used to select the model. However, since we assume a linear model (*i.e.*, absolute effects) it is natural to use a method that can estimate the absolute effects in the model selection process. We also wanted the method that was expected to have a high power, and based on previous studies, [11], ML_{LN} was expected to have higher power than e.g., WLS and LS_{lin}.

5. Conclusions

In medical research we often want to identify and quantify associations using regression analysis. Log-normal data are common and there are situations when the absolute effects are of interest (rather than the relative) and thus there is a need for linear regression methods on untransformed log-normal data. We have evaluated several regression methods using both large scale simulations of personal exposure to PM, and by applying the methods to data on biomarkers (CRP and HOMA-IR). The LS_{exp} does not provide estimates of the absolute effects and the expected outcome can be biased. The LS_{lin} and GLM_G provide correct point estimates of the expected outcome, but confidence intervals with incorrect coverage. The ML_{LN} and GLM_N worked best (unbiased estimates, narrow confidence intervals), although ML_{LN} tends to have a slightly more correct coverage for the confidence intervals.

Acknowledgments

This project was funded by the Swedish state under the agreement between the Swedish government and county councils concerning economic support for research and education of doctors (ALF-agreement).

Author Contributions

Sara Gustavsson and Eva M. Andersson were responsible for the statistical data analyses and for the manuscript. Gerd Sallsten serves as Sara Gustavsson's assistant supervisor and contributed in the modelling of exposure to particles. Björn Fagerberg is responsible for the DIWA study, and contributed with important information on diabetes, obesity and biomarkers. All authors approved the final manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Rappaport, S. Selection of the measures of exposure for epidemiology studies. *Appl. Occup. Environ. Hyg.* **1991**, *6*, 448–457.
2. Crump, K. On summarizing group exposures in risk assessment: Is an arithmetic mean or a geometric mean more appropriate? *Risk Anal.* **1998**, *18*, 293–297.
3. Rappaport, S. Assessment of long-term exposures to toxic substances in air. *Ann. Occup. Hyg.* **1991**, *35*, 61–121.

4. Koch, A. The logarithm in biology 1. Mechanisms generating the log-normal distribution exactly. *J. Theor. Biol.* **1966**, *12*, 276–290.
5. Osvoll, P.; Woldbæk, T. Distribution and skewness of occupational exposure sets of measurements in the Norwegian industry. *Ann. Occup. Hyg.* **1999**, *43*, 421–428.
6. Limpert, E.; Stahel, W.; Abbt, M. Log-normal distributions across the sciences: Keys and clues. *BioScience* **2001**, *51*, 341–352.
7. Zhou, X.-H.; Stroupe, K.; Tierney, W. Regression analysis of health care charges with heteroscedasticity. *J. R. Stat. Soc. Ser. C* **2001**, *50*, 303–312.
8. Rothman, K.J. *Epidemiology. An Introduction*; Oxford University Press Inc: New York, NY, USA, 2002.
9. Rothman, K.J.; Greenland, S. Concepts of Interaction. In *Modern Epidemiology*, 2nd ed.; Rothman, K.J., Greenland, S., Eds.; Lippincott Williams and Wilkins: Philadelphia, PA, USA, 1998.
10. McCullagh, P.; Nelder, J. *Generalized Linear Models*, 2nd ed.; CRC Press: Boca Raton, FL, USA, 1989.
11. Gustavsson, S.; Johannesson, S.; Sallsten, G.; Andersson, E.M. Linear maximum likelihood regression analysis for untransformed log-normally distributed data. *Open J. Stat.* **2012**, *2*, 389–400.
12. Yurgens, Y. *Quantifying Environmental Impact by Log-Normal Regression Modelling of Accumulated Exposure*; Chalmers University of technology and Goteborg University: Goteborg, Sweden, 2004.
13. Jensen, S.; Johansen, S.; Lauritzen, S. Globally convergent algorithms for maximizing likelihood function. *Biometrika* **1991**, *78*, 867–877.
14. Niwitpong, S. Confidence intervals for the mean of a lognormal distribution. *Appl. Math. Sci.* **2013**, *7*, 161–166.
15. Johannesson, S.; Gustafson, P.; Molnar, P.; Barregard, L.; Sallsten, G. Exposure to fine particles (PM_{2.5} and PM₁) and black smoke in the general population: Personal, indoor, and outdoor levels. *J. Expos. Sci. Environ. Epidemiol.* **2007**, *17*, 613–624.
16. Englert, N. Fine particles and human health—A review of epidemiological studies. *Toxicol. Letters* **2004**, *149*, 235–242.
17. Dominici, F.; Peng, R.D.; Bell, M.L.; Pham, L.; McDermott, A.; Zeger, S.L.; Samet, J.M. Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. *J. Am. Med. Assoc.* **2006**, *295*, 1127–1134.
18. Koistinen, K.J.; Hänninen, O.; Rotko, T.; Edwards, R.D.; Moschandreas, D.; Jantunen, M.J. Behavioral and environmental determinants of personal exposures to PM_{2.5} in EXPOLIS—Helsinki, Finland. *Atmos. Environ.* **2001**, *35*, 2473–2481.
19. Brohall, G.; Behre, C.-J.; Hulthe, J.; Wikstrand, J.; Fagerberg, B. Prevalence of diabetes and impaired glucose tolerance in 64-year-old swedish women. *Diabetes Care* **2006**, *29*, 363–367.
20. Fagerberg, B.; Kellis, D.; Bergström, G.; Behre, C.J. Adiponectin in relation to insulin sensitivity and insulin secretion in the development of type 2 diabetes: A prospective study in 64-year-old women. *J. Int. Med.* **2011**, *269*, 636–643.

21. Alberti, K.; Zimmet, P. Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: Diagnosis and classification of diabetes mellitus. Provisional report of a WHO consultation. *Diabetic Med.* **1998**, *15*, 539–553.
22. Ford, E.S. Body mass index, diabetes, and C-reactive protein among U.S. adults. *Diabetes Care* **1999**, *22*, 1971–1977.
23. Fröhlich, M.; Sund, M.; Löwel, H.; Imhof, A.; Hoffmeister, A.; Koenig, W. Independent association of various smoking characteristics with markers of systemic inflammation in men. *Eur. Heart J.* **2003**, *24*, 1365–1372.
24. Leinonen, E.; Hurt-Camejo, E.; Wiklund, O.; Hulten, L.M.; Hiukka, A.; Taskinen, M.R. Insulin resistance and adiposity correlate with acute-phase reaction and soluble cell adhesion molecules in type 2 diabetes. *Atherosclerosis* **2003**, *166*, 387–394.
25. O’Loughlin, J.; Lambert, M.; Karp, I.; McGrath, J.; Gray-Donald, K.; Barnett, T.A.; Delvin, E.E.; Levy, E.; Paradis, G. Association between cigarette smoking and C-reactive protein in a representative, population-based sample of adolescents. *Nicot. Tob. Res.* **2008**, *10*, 525–532.
26. Reaven, G. Banting lecture 1988. Role of insulin resistance in human disease. *Diabetes* **1988**, *37*, 1595–1607.
27. Sites, C.K.; Calles-Escandón, J.; Brochu, M.; Butterfield, M.; Ashikaga, T.; Poehlman, E.T. Relation of regional fat distribution to insulin sensitivity in postmenopausal women. *Fertil. Steril.* **2000**, *73*, 61–65.
28. Wagenknecht, L.E.; Langefeld, C.D.; Scherzinger, A.L.; Norris, J.M.; Haffner, S.M.; Saad, M.F.; Bergman, R.N. Insulin sensitivity, insulin secretion, and abdominal fat. The insulin resistance atherosclerosis study (IRAS) family study. *Diabetes* **2003**, *52*, 2490–2496.
29. Facchini, F.S.; Hollenbeck, C.B.; Jeppesen, J.; Chen, Y.D.; Reaven, G.M. Insulin resistance and cigarette smoking. *Lancet* **1992**, *339*, 1128–1130.
30. Mayer-Davis, E.J.; D’Agostino, R., Jr.; Karter, A.J.; Haffner, S.M.; Rewers, M.J.; Mohammed, S.; Bergman, R.N.; for the IRAS Investigators. Intensity and amount of physical activity in relation to insulin sensitivity. *J. Am. Med. Assoc.* **1998**, *279*, 669–674.
31. Matthews, D.R.; Hosker, J.P.; Rudenski, A.S.; Naylor, B.A.; Treacher, D.F.; Turner, R.C. Homeostasis model assessment: Insulin resistance and β -cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia* **1985**, *28*, 412–419.
32. Taniguchi, A.; Fukushima, M.; Sakai, M.; Kataoka, K.; Nagata, I.; Doi, K.; Arakawa, H.; Nagasaka, S.; Toshikatsu, K.; Nakai, Y. The role of the body mass index and triglyceride levels in identifying insulin-sensitive and insulin-resistant variants in Japanese non-insulin-dependent diabetic patients. *Metabolism* **2000**, *49*, 1001–1005.
33. Radikova, Z.; Koska, J.; Huckova, M.; Ksinantova, L.; Imrich, R.; Trnovec, T.; Langer, P.; Sebkova, E.; Klimes, I. Insulin sensitivity indices: A proposal of cut-off points for simple identification of insulin-resistant subjects. *Exp. Clin. Endocrinol. Diabetes* **2006**, *114*, 249–256.
34. Geloneze, B.; Vasques, A.C.J.; Stabe, C.F.C.; Pareja, J.C.; de Lima Rosado, L.E.F.P.; de Queiroz, E.C.; Tambascia, M.A.; BRAMS Investigators. HOMA1-IR and HOMA2-IR indexes in identifying insulin resistance and metabolic syndrome: Brazilian metabolic syndrome study (BRAMS). *Arq. Bra. Endocrinol. Metab.* **2009**, *53*, 281–287.

35. Dickerson, E.H.; Cho, L.W.; Maguiness, S.D.; Killick, S.L.; Atkin, S.L. Insulin resistance and free androgen index correlate with the outcome of controlled ovarian hyperstimulation in non-PCOS women undergoing IVF. *Hum. Reprod.* **2010**, *25*, 504–509.
36. Land, C.E. An evaluation of approximate confidence interval estimation methods for lognormal means. *Technometrics* **1972**, *14*, 145–158.
37. Zhou, X.-H.; Gao, S.; Hui, S. Methods for comparing the means of two independent log-normal samples. *Biometrics* **1997**, *53*, 1129–1135.
38. Zou, G.Y.; Huo, C.Y.; Taleban, J. Simple confidence intervals for lognormal means and their differences with environmental applications. *Environmetrics* **2009**, *20*, 172–180.
39. Taylor, D.J.; Kupper, L.L.; Muller, K.E. Improved approximate confidence intervals for the mean of a log-normal random variable. *Stat. Med.* **2002**, *21*, 1443–1459.
40. Wu, J.; Wong, A.C.M.; Jiang, G. Likelihood-based confidence intervals for a log-normal mean. *Stat. Med.* **2003**, *22*, 1849–1860.
41. Firth, D. Multiplicative errors: Log-normal or gamma? *J. R. Stat. Soc. Ser. B* **1998**, *50*, 266–268.
42. Das, R.N.; Park, J.-S. Discrepancy in regression estimates between log-normal and gamma: Some case studies. *J. Appl. Stat.* **2012**, *39*, 97–111.
43. Rade, L.; Westergran, B. *Mathematics Handbook for Science and Engineering (BETA)*; Studentlitteratur: Lund, Sweden, 1998.
44. Visser, M.; Bouter, L.M.; McQuillan, G.M.; Wener, M.H.; Harris, T.B. Elevated C-reactive protein levels in overweight and obese adults. *J. Am. Med. Assoc.* **1999**, *282*, 2131–2135.
45. Yudkin, J.S.; Stehouwer, C.D.A.; Emeis, J.J.; Coppack, S.W. C-Reactive protein in healthy subjects: Associations with obesity, insulin resistance, and endothelial dysfunction. A potential role for cytokines originating from adipose tissue? *Arterioscler. Thromb. Vasc. Biol.* **1999**, *19*, 972–978.
46. Pannacciulli, N.; Cantatore, F.P.; Minenna, A.; Bellacicco, M.; Giorgino, R.; de Pergola, G. C-reactive protein is independently associated with total body fat, central fat, and insulin resistance in adult women. *Int. J. Obes. Relat. Metab. Disord.* **2001**, *25*, 1416–1420.
47. McLaughlin, T.; Abbasi, F.; Lamendola, C.; Liang, L.; Reaven, G.; Schaaf, P.; Reaven, P. Differentiation between obesity and insulin resistance in the association with C-reactive protein. *Circulation* **2002**, *106*, 2908–2912.
48. Lapice, E.; Maione, S.; Patti, L.; Cipriano, P.; Rivellese, A.A.; Riccardi, G.; Vaccaro, O. Abdominal adiposity is associated with elevated C-reactive protein independent of bmi in healthy nonobese people. *Diabetes Care* **2009**, *32*, 1734–1736.
49. Brooks, G.; Blaha, M.; Blumenthal, R. Relation of C-reactive protein to abdominal adiposity. *Am. J. Cardiol.* **2010**, *106*, 56–61.
50. Hermsdorff, H.H.M.; Zulet, M.A.; Puchau, B.; Martinez, J.A. Central adiposity rather than total adiposity measurements are specifically involved in the inflammatory status from healthy young adults. *Inflammation* **2011**, *34*, 161–170.
51. Hak, A.E.; Stehouwer, C.D.A.; Bots, M.L.; Polderman, K.H.; Schalkwijk, C.G.; Westendorp, I.C.D.; Hofman, A.; Witteman, J.C.M. Associations of C-reactive protein with measures of obesity, insulin resistance, and subclinical atherosclerosis in healthy, middle-aged women. *Arterioscler. Thromb. Vasc. Biol.* **1999**, *19*, 1986–1991.

52. Festa, A.; D'Agostino, R., Jr.; Howard, G.; Mykkanen, L.; Tracy, R.P.; Haffner, S.M. Chronic subclinical inflammation as part of the insulin resistance syndrome: The insulin resistance atherosclerosis study (IRAS). *Circulation* **2000**, *102*, 42–47.
53. Lemieux, I.; Pascot, A.; Prud'homme, D.; Almeras, N.; Bogaty, P.; Nadeau, A.; Bergeron, J.; Despres, J.-P. Elevated C-reactive protein : Another component of the atherothrombotic profile of abdominal obesity. *Arterioscler. Thromb. Vasc. Biol.* **2001**, *21*, 961–967.
54. Wallace, T.M.; Levy, J.; Matthews, D. Use and abuse of HOMA modeling. *Diabetes Care* **2004**, *27*, 1487–1495.
55. Huang, L.-H.; Liao, Y.-L.; Hsu, C.-H. Waist circumference is a better predictor than body mass index of insulin resistance in type 2 diabetes. *Obes. Res. Clin. Pract.* **2011**, *6*, e314–e320.
56. Lee, K. Usefulness of the metabolic syndrome criteria as predictors of insulin resistance among obese Korean women. *Public Health Nutr.* **2010**, *13*, 181–186.
57. Thomas, D.C. General relative-risk models for survival time and matched case-control analysis. *Biometrics* **1981**, *37*, 673–686.
58. Richardson, D.B.; Langholz, B. Background stratified poisson regression analysis of cohort data. *Radiat. Environ. Biophys.* **2012**, *51*, 15–22.
59. Rothman, K.J. Causes. *Am. J. Epidemiol.* **1976**, *104*, 587–592.
60. VanderWeele, T.J. On the distinction between interaction and effect modification. *Epidemiology* **2009**, *20*, 863–871.
61. Nurminen, M. To use or not to use the odds ratio in epidemiologic analyses? *Eur. J. Epidemiol.* **1995**, *11*, 365–371.
62. Andersson, T.; Alfredsson, L.; Kallberg, H.; Zdravkovic, S.; Ahlbom, A. Calculating measures of biological interaction. *Eur. J. Epidemiol.* **2005**, *20*, 575–579.