Environ Health Perspect

DOI: 10.1289/EHP14784

Note to readers with disabilities: *EHP* strives to ensure that all journal content is accessible to all readers. However, some figures and Supplemental Material published in *EHP* articles may not conform to 508 standards due to the complexity of the information being presented. If you need assistance accessing journal content, please contact ehpsubmissions@niehs.nih.gov. Our staff will work with you to assess and meet your accessibility needs within 3 working days.

Supplemental Material

Calibration of Toenail Metal Concentrations for Sample Mass Heterogeneity and Between-Batch Variability: The COMET Approach

Roberto Pastor-Barriuso, Enrique Gutiérrez-González, Elena Varea-Jiménez, José Luis Gómez-Ariza, Gemma Castaño-Vinyals, Nuria Aragonés, Antonio José Molina, Trinidad Dierssen-Sotos, Guillermo Fernández-Tardón, Pilar Amiano, María Ederra-Sanz, Victor Moreno, José Juan Jiménez-Moleón, Ana Molina-Barceló, Rafael Marcos-Gragera, Delphine Casabonne, Juan Alguacil, Jesús Humberto Gómez-Gómez, Tamara García-Barrera, Manolis Kogevinas, Marina Pollán, and Beatriz Pérez-Gómez

Table of Contents

Figure S1. Measured metal concentrations in homogenized toenail samples of different mass in the reproducibility study (*n* = 29). Results were based on a pooled sample of toenails, which was cryohomogenized and the resulting powder divided into ten samples of 5 and 10 mg and three samples of 30, 50, and 100 mg. Note: ppb, parts per billion; ppm, parts per million.

Figure S2. Geometric mean metal concentrations as a smooth function of toenail sample mass within each laboratory batch in a population-based multicase-control study in Spain, 2008–2013 (*n* = 7,923). Batch-specific trends in geometric mean metal concentrations (gray curves) were obtained from separate regression models within each laboratory batch relating log-transformed metal concentrations with a natural cubic spline of log-transformed toenail sample mass with internal knots at the overall 35th and 65th percentiles and boundary knots at the 5th and 95th percentiles (3.0, 13.9, 28.9, and 63.2 mg, respectively). Batch-specific analyses were restricted to 119 of the 122 laboratory batches (97.5%) with at least 25 metal determinations. Note: ppb, parts per billion.

Figure S3. Batch compositional bias in measured metal concentrations associated with toenail samples of the same mass analyzed in laboratory batches with different geometric mean masses in a population-based multicase-control study in Spain, 2008–2013 (*n* = 7,923). The batch compositional bias in measured metal concentrations (thick curve) and its 95% confidence interval (thin curves) were determined by the estimates and standard errors of the fixed spline parameters η_1, η_2, and η_3 associated with mean log-transformed masses for each laboratory batch in heteroscedastic spline mixed models for log-transformed metal concentrations. The reference level was set at the geometric mean mass over all laboratory batches (17.6 mg).

References

Development of the Heteroscedastic Spline Mixed Model

To correct the measured metal concentrations for both the heterogeneity in toenail sample masses and the variability between laboratory batches, we developed a spline mixed model with heteroscedastic errors by using a two-level hierarchical approach.[1,2] At the first within-batch level, the log-transformed measured metal concentration Y_{ij} for nail specimen $j = 1,\ldots, n_i$ in laboratory batch $i = 1,\ldots, I$ was assumed to be related to the log-transformed mass of the nail clipping m_{ij} for that specimen, the case-control status $\mathbf{z}_{1ij}$ of the corresponding participant, and the participant's sociodemographic factors $\mathbf{z}_{2ij}$ through the spline regression model

$$Y_{ij} = \alpha_{0i} + \boldsymbol{\delta}_i'\{\mathbf{s}(m_{ij}) - \mathbf{s}(\overline{m})\} + \boldsymbol{\alpha}_{1i}'\mathbf{z}_{1ij} + \boldsymbol{\alpha}_{2i}'\mathbf{z}_{2ij} + \varepsilon_{ij},$$

where the intercept α_{0i} was the mean log-transformed metal concentration for laboratory batch i at the overall mean log-transformed nail clipping mass $\overline{m}$ for controls with reference levels of sociodemographic factors and the parameters $\boldsymbol{\delta}_i = (\delta_{1i}, \delta_{2i}, \delta_{3i})'$ for the B-spline basis functions $\mathbf{s}(m_{ij}) = (s_1(m_{ij}), s_2(m_{ij}), s_3(m_{ij}))'$ determined the systematic bias in the log-transformed measured metal concentrations for that batch as a natural cubic spline of m_{ij} with internal knots at the 35th and 65th percentiles and boundary knots at the 5th and 95th percentiles of the overall log-transformed mass distribution.[3] The parameters $\boldsymbol{\alpha}_{1i}$ were the mean shifts in log-transformed metal concentrations for batch i between cases of each tumor type and disease-free controls and $\boldsymbol{\alpha}_{2i}$ were the mean differences for that batch comparing each level of sociodemographic factors with the reference one. The within-batch errors ε_{ij} were assumed to be independent and normally distributed with mean 0 and heterogeneous variance σ_{ij}^2 of the log-linear form

$$\log(\sigma_{ij}) = \phi_i + \boldsymbol{\gamma}_i'\{\mathbf{s}(m_{ij}) - \mathbf{s}(\overline{m})\},$$

which was also related to the log-transformed mass of the nail clipping through a natural cubic spline with the same internal and boundary knots described above. The scale $\exp(2\phi_i)$ was the variance of the log-transformed metal concentrations for batch i at the overall mean log-transformed mass $\overline{m}$ and the parameters $\boldsymbol{\gamma}_i = (\gamma_{1i}, \gamma_{2i}, \gamma_{3i})'$ determined the variance ratio for that batch as the exponential of a natural cubic spline of the log-transformed mass.[2]

The second level represented the variation in model coefficients across laboratory batches. The batch-specific intercepts α_{0i} were related to the mean log-transformed nail clipping masses for each batch $\overline{m}_i$ through a natural cubic spline with the same previous knots, thus allowing for different systematic biases in the measured metal concentrations for toenail samples of the same mass analyzed in batches with different geometric mean masses.[1] The spline parameters that determined the systematic bias across toenail samples of different mass $\boldsymbol{\delta}_i$ were allowed to vary randomly across batches, whereas the mean differences in log-transformed metal concentrations by case-control status $\boldsymbol{\alpha}_{1i}$ and sociodemographic factors $\boldsymbol{\alpha}_{2i}$, together with the variance parameters ϕ_i and $\boldsymbol{\gamma}_i$, were assumed to be fixed for all batches. Thus, the model at the second between-batch level was

$$\alpha_{0i} = \alpha_0 + \boldsymbol{\eta}'\{\mathbf{s}(\overline{m}_i) - \mathbf{s}(\overline{m})\} + a_i,$$

$$\boldsymbol{\delta}_i = \boldsymbol{\delta} + \boldsymbol{d}_i,$$

$$\boldsymbol{\alpha}_{1i} = \boldsymbol{\alpha}_1,\ \boldsymbol{\alpha}_{2i} = \boldsymbol{\alpha}_2,\ \phi_i = \phi,\ \boldsymbol{\gamma}_i = \boldsymbol{\gamma},$$

where the random between-batch variations a_i and $\boldsymbol{d}_i = (d_{1i}, d_{2i}, d_{3i})'$ were assumed to be independent of the within-batch errors ε_{ij} and to follow a multivariate normal distribution with mean $\mathbf{0}$ and unstructured variance-covariance matrix $\boldsymbol{\Sigma}$. Combining the two nested models yielded the spline mixed model

$$Y_{ij} = \alpha_0 + a_i + (\boldsymbol{\delta} + \boldsymbol{d}_i)'\{\mathbf{s}(m_{ij}) - \mathbf{s}(\overline{m})\}$$

$$+ \boldsymbol{\eta}'\{\mathbf{s}(\overline{m}_i) - \mathbf{s}(\overline{m})\} + \boldsymbol{\alpha}_1'\mathbf{z}_{1ij} + \boldsymbol{\alpha}_2'\mathbf{z}_{2ij} + \varepsilon_{ij},$$

with heterogeneous within-batch variance

$$\log(\sigma_{ij}) = \phi + \boldsymbol{\gamma}'\{\mathbf{s}(m_{ij}) - \mathbf{s}(\overline{m})\},$$

which accounted for the effects of the toenail sample mass and the laboratory batch on the measured metal concentration.

Variance Components in Heteroscedastic Linear Mixed Models

In this section, we derive a general expression for the variance components in linear mixed models with heteroscedastic within-group errors, which was used to decompose the total variance of measured metal concentrations in the present paper. The heteroscedastic linear mixed model for the response Y_{ij} of observation j in group i can be expressed as

$$Y_{ij} = \boldsymbol{\beta}'\mathbf{x}_{ij} + \boldsymbol{b}_i'\mathbf{z}_{ij} + \varepsilon_{ij},$$

where $\boldsymbol{\beta}$ is the vector of fixed effects associated with covariates $\mathbf{x}_{ij}$, $\boldsymbol{b}_i$ is the vector of random effects for covariates $\mathbf{z}_{ij}$, which varies across groups with mean $\mathbf{0}$ and variance-covariance matrix $\boldsymbol{\Sigma}$, and ε_{ij} is the within-group error with mean 0 and heterogeneous variance of the log-linear form $\sigma_{ij}^2 = \exp(2\boldsymbol{\theta}'\mathbf{v}_{ij})$, with the vector of variance parameters $\boldsymbol{\theta}$ associated with covariates $\mathbf{v}_{ij}$.[2] The random effects $\boldsymbol{b}_i$ and the within-group errors ε_{ij} are assumed independent, but no requirements are needed about their distributional forms.

The total variance of the response Y_{ij} can be decomposed into its conditional mean and variance given covariates $\mathbf{x}_{ij}$, $\mathbf{z}_{ij}$, and $\mathbf{v}_{ij}$ as

$$\begin{aligned}\mathrm{var}(Y_{ij}) &= \mathrm{var}\{E(Y_{ij}|\mathbf{x}_{ij}, \mathbf{z}_{ij}, \mathbf{v}_{ij})\} + E\{\mathrm{var}(Y_{ij}|\mathbf{x}_{ij}, \mathbf{z}_{ij}, \mathbf{v}_{ij})\} \\ &= \mathrm{var}(\boldsymbol{\beta}'\mathbf{x}_{ij}) + E\{\mathbf{z}_{ij}'\boldsymbol{\Sigma}\mathbf{z}_{ij} + \exp(2\boldsymbol{\theta}'\mathbf{v}_{ij})\},\end{aligned}$$

since $\boldsymbol{b}_i$ and ε_{ij} are independent. If $\boldsymbol{\mu}_x = E(\mathbf{x}_{ij})$ and $\mathbf{V}_x = \mathrm{var}(\mathbf{x}_{ij})$ are the mean vector and variance-covariance matrix for fixed-effects covariates $\mathbf{x}_{ij}$, $\boldsymbol{\mu}_z = E(\mathbf{z}_{ij})$ and $\mathbf{V}_z = \mathrm{var}(\mathbf{z}_{ij})$ for random-effects covariates $\mathbf{z}_{ij}$, and $\boldsymbol{\mu}_v = E(\mathbf{v}_{ij})$ and $\mathbf{V}_v = \mathrm{var}(\mathbf{v}_{ij})$ for variance covariates $\mathbf{v}_{ij}$, standard matrix algebra[4] can be applied to obtain

$$\begin{aligned}\mathrm{var}(Y_{ij}) &= \boldsymbol{\beta}'\mathbf{V}_x\boldsymbol{\beta} + \mathrm{tr}(\boldsymbol{\Sigma}\mathbf{V}_z) + \boldsymbol{\mu}_z'\boldsymbol{\Sigma}\boldsymbol{\mu}_z + E\{\exp(2\boldsymbol{\theta}'\mathbf{v}_{ij})\} \\ &\approx \boldsymbol{\beta}'\mathbf{V}_x\boldsymbol{\beta} + \mathrm{tr}(\boldsymbol{\Sigma}\mathbf{V}_z) + \boldsymbol{\mu}_z'\boldsymbol{\Sigma}\boldsymbol{\mu}_z + \exp\{2E(\boldsymbol{\theta}'\mathbf{v}_{ij})\} + 2\exp\{2E(\boldsymbol{\theta}'\mathbf{v}_{ij})\}\mathrm{var}(\boldsymbol{\theta}'\mathbf{v}_{ij}) \\ &= \boldsymbol{\beta}'\mathbf{V}_x\boldsymbol{\beta} + \mathrm{tr}(\boldsymbol{\Sigma}\mathbf{V}_z) + \boldsymbol{\mu}_z'\boldsymbol{\Sigma}\boldsymbol{\mu}_z + \exp(2\boldsymbol{\theta}'\boldsymbol{\mu}_v) + 2\exp(2\boldsymbol{\theta}'\boldsymbol{\mu}_v)\boldsymbol{\theta}'\mathbf{V}_v\boldsymbol{\theta},\end{aligned}$$

where the matrix trace $\mathrm{tr}(\boldsymbol{\Sigma}\mathbf{V}_z)$ is the sum of its main diagonal elements and the approximation is derived from the second-order Taylor expansion of the exponential function about $2\boldsymbol{\theta}'\boldsymbol{\mu}_v$. Thus, the total variance of the response is partitioned into the variance explained by fixed-effects covariates $\boldsymbol{\beta}'\mathbf{V}_x\boldsymbol{\beta}$, the between-group variance due to variation in random-effects covariates $\mathrm{tr}(\boldsymbol{\Sigma}\mathbf{V}_z)$, the between-group variance at mean values of random-effects covariates $\boldsymbol{\mu}_z'\boldsymbol{\Sigma}\boldsymbol{\mu}_z$, the within-group variance explained by variance covariates $2\exp(2\boldsymbol{\theta}'\boldsymbol{\mu}_v)\boldsymbol{\theta}'\mathbf{V}_v\boldsymbol{\theta}$, and the within-group residual variance $\exp(2\boldsymbol{\theta}'\boldsymbol{\mu}_v)$. This partition extends the variance components in standard linear mixed models to the presence of heteroscedastic within-group errors.[5]

Given a subset $\mathbf{x}_{1ij}$ of the fixed-effects covariates $\mathbf{x}_{ij} = (\mathbf{x}_{1ij}', \mathbf{x}_{2ij}')'$ not included in the random-effects and variance terms of the heteroscedastic linear mixed model, the conditional variance of the response Y_{ij} can be decomposed in a similar way as

$$\mathrm{var}(Y_{ij}|\mathbf{x}_{1ij}) \approx \boldsymbol{\beta}_2'\mathbf{V}_{x_2|x_1}\boldsymbol{\beta}_2 + \mathrm{tr}(\boldsymbol{\Sigma}\mathbf{V}_{z|x_1}) + \boldsymbol{\mu}_{z|x_1}'\boldsymbol{\Sigma}\boldsymbol{\mu}_{z|x_1}$$
$$+ \exp(2\boldsymbol{\theta}'\boldsymbol{\mu}_{v|x_1}) + 2\exp(2\boldsymbol{\theta}'\boldsymbol{\mu}_{v|x_1})\boldsymbol{\theta}'\mathbf{V}_{v|x_1}\boldsymbol{\theta},$$

where $\boldsymbol{\beta}_2$ is the vector of fixed effects associated with covariates $\mathbf{x}_{2ij}$, $\boldsymbol{\mu}_{x_2|x_1} = E(\mathbf{x}_{2ij}|\mathbf{x}_{1ij})$, $\boldsymbol{\mu}_{z|x_1} = E(\mathbf{z}_{ij}|\mathbf{x}_{1ij})$, and $\boldsymbol{\mu}_{v|x_1} = E(\mathbf{v}_{ij}|\mathbf{x}_{1ij})$ are the conditional mean vectors for $\mathbf{x}_{2ij}$, $\mathbf{z}_{ij}$, and $\mathbf{v}_{ij}$ given $\mathbf{x}_{1ij}$, and $\mathbf{V}_{x_2|x_1} = \mathrm{var}(\mathbf{x}_{2ij}|\mathbf{x}_{1ij})$, $\mathbf{V}_{z|x_1} = \mathrm{var}(\mathbf{z}_{ij}|\mathbf{x}_{1ij})$, and $\mathbf{V}_{v|x_1} = \mathrm{var}(\mathbf{v}_{ij}|\mathbf{x}_{1ij})$ are their conditional variance-covariance matrices. Closed-form expressions for these conditional means and variances can be derived under the standard assumptions of linearity and homogeneity of variance. If the conditional mean of $\mathbf{x}_{2ij}$ given $\mathbf{x}_{1ij}$ is linear $\boldsymbol{\mu}_{x_2|x_1} = \mathbf{a} + \mathbf{B}\mathbf{x}_{1ij}$ and its conditional variance is constant $\mathbf{V}_{x_2|x_1} = \mathbf{C}$, it follows that

$$\boldsymbol{\mu}_{x_2} = E(\mathbf{x}_{2ij}) = E\{E(\mathbf{x}_{2ij}|\mathbf{x}_{1ij})\} = \mathbf{a} + \mathbf{B}E(\mathbf{x}_{1ij}) = \mathbf{a} + \mathbf{B}\boldsymbol{\mu}_{x_1},$$

$$\mathbf{V}_{x_2} = \mathrm{var}(\mathbf{x}_{2ij}) = \mathrm{var}\{E(\mathbf{x}_{2ij}|\mathbf{x}_{1ij})\} + E\{\mathrm{var}(\mathbf{x}_{2ij}|\mathbf{x}_{1ij})\}$$
$$= \mathrm{var}(\mathbf{a} + \mathbf{B}\mathbf{x}_{1ij}) + \mathbf{C} = \mathbf{B}\mathrm{var}(\mathbf{x}_{1ij})\mathbf{B}' + \mathbf{C} = \mathbf{B}\mathbf{V}_{x_1}\mathbf{B}' + \mathbf{C},$$

$$\mathbf{V}_{x_1x_2} = \text{cov}(\mathbf{x}_{1ij}, \mathbf{x}_{2ij}) = \text{cov}\{E(\mathbf{x}_{1ij}|\mathbf{x}_{1ij}), E(\mathbf{x}_{2ij}|\mathbf{x}_{1ij})\} + E\{\text{cov}(\mathbf{x}_{1ij}, \mathbf{x}_{2ij}|\mathbf{x}_{1ij})\}$$

$$= \text{cov}(\mathbf{x}_{1ij}, \mathbf{a} + \mathbf{B}\mathbf{x}_{1ij}) = \text{var}(\mathbf{x}_{1ij})\mathbf{B}' = \mathbf{V}_{x_1}\mathbf{B}',$$

so that

$$\boldsymbol{\mu}_{x_2|x_1} = \boldsymbol{\mu}_{x_2} + \mathbf{B}(\mathbf{x}_{1ij} - \boldsymbol{\mu}_{x_1}) = \boldsymbol{\mu}_{x_2} + \mathbf{V}_{x_1x_2}{}'\mathbf{V}_{x_1}^{-1}(\mathbf{x}_{1ij} - \boldsymbol{\mu}_{x_1}),$$

$$\mathbf{V}_{x_2|x_1} = \mathbf{V}_{x_2} - \mathbf{B}\mathbf{V}_{x_1}\mathbf{B}' = \mathbf{V}_{x_2} - \mathbf{V}_{x_1x_2}{}'\mathbf{V}_{x_1}^{-1}\mathbf{V}_{x_1x_2}.$$

Analogous expressions can readily be derived for the conditional mean vectors and variance-covariance matrices of $\mathbf{z}_{ij}$ and $\mathbf{v}_{ij}$ given $\mathbf{x}_{1ij}$. Note that these results hold for any arbitrary distributional forms of the fixed-effects, random-effects, and variance covariates.

The total and conditional variances of the response can be estimated from the above expressions by replacing the fixed effects $\boldsymbol{\beta}$, the variance-covariance matrix $\boldsymbol{\Sigma}$ of random effects, and the variance parameters $\boldsymbol{\theta}$ by their estimates $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\Sigma}}$, and $\hat{\boldsymbol{\theta}}$ from the heteroscedastic mixed model, and the marginal expectations $\boldsymbol{\mu}$ and variance-covariance matrices $\mathbf{V}$ of the fixed-effects, random-effects, and variance covariates by their sample estimates $\bar{\mathbf{x}}$ and $\mathbf{S}$. Thus, the estimated conditional variance of the response Y_{ij} given $\mathbf{x}_{1ij}$ at its sample mean $\bar{\mathbf{x}}_1$ becomes

$$\widehat{\text{var}}(Y_{ij}|\bar{\mathbf{x}}_1) \approx \hat{\boldsymbol{\beta}}_2{}'\mathbf{S}_{x_2|x_1}\hat{\boldsymbol{\beta}}_2 + \text{tr}(\hat{\boldsymbol{\Sigma}}\mathbf{S}_{z|x_1}) + \bar{\mathbf{z}}'\hat{\boldsymbol{\Sigma}}\,\bar{\mathbf{z}}$$

$$+ \exp(2\hat{\boldsymbol{\theta}}\bar{\mathbf{v}}) + 2\exp(2\hat{\boldsymbol{\theta}}\bar{\mathbf{v}})\hat{\boldsymbol{\theta}}\mathbf{S}_{v|x_1}\hat{\boldsymbol{\theta}},$$

since the estimated conditional means of $\mathbf{x}_{2ij}$, $\mathbf{z}_{ij}$, and $\mathbf{v}_{ij}$ at $\bar{\mathbf{x}}_1$ reduce to

$$\hat{\boldsymbol{\mu}}_{x_2|x_1}(\bar{\mathbf{x}}_1) = \bar{\mathbf{x}}_2 + \mathbf{S}_{x_1x_2}{}'\mathbf{S}_{x_1}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_1) = \bar{\mathbf{x}}_2,$$

$\hat{\boldsymbol{\mu}}_{z|x_1}(\bar{\mathbf{x}}_1) = \bar{\mathbf{z}}$, and $\hat{\boldsymbol{\mu}}_{v|x_1}(\bar{\mathbf{x}}_1) = \bar{\mathbf{v}}$, and their estimated conditional variances

$$\mathbf{S}_{x_2|x_1} = \mathbf{S}_{x_2} - \mathbf{S}_{x_1x_2}{}'\mathbf{S}_{x_1}^{-1}\mathbf{S}_{x_1x_2},$$

$\mathbf{S}_{z|x_1}$, and $\mathbf{S}_{v|x_1}$ are constant.

R Script COMET

This section provides the annotated R script COMET (COrrected METals) to fit the heteroscedastic spline mixed model, extract variance components, and calibrate metal concentrations.

Step 1. Data management. The data frame `data` contains one row per participant with the measured metal concentration in numeric variable `metal`, the toenail sample mass in numeric variable `mass`, the laboratory batch in factor `batch`, the case-control status in factor `cancer`, and the sociodemographic characteristics in factors `province`, `sex`, `age.cat`, and `educ`. We first compute the individual log-transformed mass `lnmass` and the mean log-transformed mass for each batch `lnmass.m`. Using function `ns` from package `splines`, we then generate the B-spline basis terms `lnmass.ns1-3` and `lnmass.m.ns1-3` for natural cubic splines of `lnmass` and `lnmass.m` with internal and boundary knots at the overall 5th, 35th, 65th, and 95th percentiles and constrained to be zero at the overall mean log-transformed mass.

```
data$lnmass <- log(data$mass)
data <- merge(data,aggregate(lnmass~batch,data=data,mean),
              by="batch",suffixes=c("",".m"))

library(splines)
knots <- quantile(data$lnmass,probs=c(0.05,0.35,0.65,0.95))
basis <- ns(data$lnmass,knots=knots[2:3],intercept=F,
            Boundary.knots=knots[c(1,4)])
data[,paste("lnmass.ns",1:3,sep="")] <-
   sweep(basis,2,predict(basis,mean(data$lnmass)))
data[,paste("lnmass.m.ns",1:3,sep="")] <-
   sweep(predict(basis,data$lnmass.m),2,predict(basis,mean(data$lnmass)))
rm(knots,basis)
```

Step 2. Heteroscedastic spline mixed model. Using function `lme` from package `nlme`, we fit a mixed model for log-transformed metal concentrations with fixed effects for mass-related spline terms `lnmass.ns1-3` and `lnmass.m.ns1-3`, case-control status, and sociodemographic factors, random between-batch effects for spline terms `lnmass.ns1-3` with unstructured variance-covariance matrix, and heterogeneous within-batch variance loglinearly related to spline terms `lnmass.ns1-3` through restricted maximum likelihood methods. We then extract the estimated fixed effects `fe.est`, the estimated variance-covariance matrix of random effects `re.var`, and the estimated variance parameters `lnsigma.est` associated with spline terms (including intercept).

```
library(nlme)
fit <- lme(log(metal)~1+lnmass.ns1+lnmass.ns2+lnmass.ns3+
                      lnmass.m.ns1+lnmass.m.ns2+lnmass.m.ns3+
                      cancer+province+sex+age.cat+educ,
           random=~1+lnmass.ns1+lnmass.ns2+lnmass.ns3|batch,
           weights=varComb(varExp(form=~lnmass.ns1),
                           varExp(form=~lnmass.ns2),
                           varExp(form=~lnmass.ns3)),
           data=data,method="REML",na.action=na.exclude,
           control=list(maxIter=250,msMaxIter=250,
                        msMaxEval=400,returnObject=T))

fe.est <- fixef(fit)[1:7]
re.var <- diag(4)
re.var[lower.tri(re.var)] <-
   na.exclude(as.numeric(VarCorr(fit,rdig=7)[2:4,3:5]))
re.var[upper.tri(re.var)] <- t(re.var)[upper.tri(re.var)]
sd <- as.numeric(VarCorr(fit)[1:4,2])
re.var <- diag(sd)%*%re.var%*%diag(sd)
dimnames(re.var) <- list(names(fixef(fit))[1:4],names(fixef(fit))[1:4])
lnsigma.est <- c(log(fit$sigma),coef(fit$modelStruct$varStruct))
names(lnsigma.est) <- names(fixef(fit))[1:4]
rm(sd)
```

Step 3. Variance components. We first estimate the conditional variance-covariance matrices of mass-related spline terms `lnmass.ns1-3` and `lnmass.m.ns1-3` given case-control status and sociodemographic factors by using their marginal variance-covariance matrices in the study sample. Combining these conditional variance-covariance matrices of spline terms with the estimated model coefficients, we then decompose the conditional variance of log-transformed measured metal concentrations given case-control status and sociodemographic factors into the variance explained by the average mass-related bias over all batches `var.comp[1]`, the between-batch variance in mass-related biases `var.comp[2]`, the between-batch variance at the mean log-transformed mass `var.comp[3]`, the within-batch variance explained by toenail sample mass `var.comp[4]`, and the within-batch residual variance at the mean log-transformed mass `var.comp[5]`.

```
Mx <- model.matrix(~1+lnmass.ns1+lnmass.ns2+lnmass.ns3+
                    lnmass.m.ns1+lnmass.m.ns2+lnmass.m.ns3,data=data)
Mz <- model.matrix(~cancer+province+sex+age.cat+educ,data=data)[,-1]
Sx <- var(Mx)
Sz <- var(Mz)
Sxz <- cov(Mx,Mz)
Sx.z <- Sx - Sxz%*%solve(Sz)%*%t(Sxz)
Sx1.z <- Sx.z[1:4,1:4]
mx1 <- apply(Mx,2,mean)[1:4]
rm(Mx,Mz,Sx,Sz,Sxz)

var.comp <- NULL
var.comp[1] <- as.numeric(fe.est%*%Sx.z%*%fe.est)
var.comp[2] <- sum(diag(re.var%*%Sx1.z))
var.comp[3] <- as.numeric(mx1%*%re.var%*%mx1)
var.comp[5] <- exp(2*as.numeric(lnsigma.est%*%mx1))
var.comp[4] <- 2*var.comp[5]*as.numeric(lnsigma.est%*%Sx1.z%*%lnsigma.est)
names(var.comp) <- c("avg.mass","bvar.mass","bvar.intcp",
                     "wvar.mass","wvar.resid")
rm(Sx.z,Sx1.z,mx1)
```

Step 4. Calibrated metal concentrations. Using the heteroscedastic spline mixed model, we calculate the log-transformed calibrated metal concentrations as the estimated marginal mean over all batches at the mean log-transformed mass `mu` plus the within-batch residuals rescaled to their estimated variance at the mean log-transformed mass `res`. These calibrated metal concentrations represent the values that would have been obtained had all toenail samples been analyzed in the same average laboratory batch and sample masses been set to the geometric mean for all participants, conditional on their case-control status and sociodemographic factors.

```
mat <- cbind(data$lnmass.ns1,data$lnmass.ns2,data$lnmass.ns3,
             data$lnmass.m.ns1,data$lnmass.m.ns2,data$lnmass.m.ns3)
mu <- fitted(fit,level=0) - as.numeric(mat%*%fe.est[-1])
res <- exp(-as.numeric(mat[,1:3]%*%lnsigma.est[-1]))*resid(fit)
data$metal.calbr <- exp(mu + res)
rm(mat,mu,res)
```

Table S1. Limits of detection for toenail metals in a population-based multicase-control study in Spain, 2008–2013 (n = 7,923).

Metal	**Limit of detection (ppb)**	**No. of participants below limit of detection (%)**	
		Cases	**Controls**
Aluminum	0.21	1 (0.0)	0 (0.0)
Vanadium	0.05	0 (0.0)	0 (0.0)
Chromium	0.05	2 (0.0)	2 (0.1)
Manganese	0.03	5 (0.1)	2 (0.1)
Iron	0.14	4 (0.1)	4 (0.1)
Cobalt	0.02	24 (0.5)	8 (0.2)
Nickel	0.02	0 (0.0)	1 (0.0)
Copper	0.04	10 (0.2)	9 (0.3)
Zinc	0.27	1 (0.0)	1 (0.0)
Arsenic	0.04	27 (0.6)	12 (0.3)
Selenium	0.05	4 (0.1)	2 (0.1)
Molybdenum	0.03	5 (0.1)	3 (0.1)
Cadmium	0.05	26 (0.6)	16 (0.5)
Thallium	0.03	14 (0.3)	12 (0.3)
Lead	0.02	16 (0.4)	9 (0.3)
Uranium	0.02	32 (0.7)	25 (0.7)

Note: ppb, parts per billion.

Table S2. Odds ratios for common cancers per two-fold increase in measured toenail metal concentrations in a population-based multicase-control study in Spain, 2008–2013 (*n* = 7,923).

Metal	**Colorectal**	**Breast**	**Prostate**	**Stomach**	**Leukemia**
No. of cases/controls[a]	1,422/3,373	1,438/1,679	915/1,273	343/2,728	355/1,524
Aluminum					
Odds ratio[b] (95% CI)	0.99 (0.94, 1.04)	1.04 (0.97, 1.11)	1.04 (0.97, 1.11)	0.84 (0.77, 0.92)	0.96 (0.87, 1.06)
Odds ratio[c] (95% CI)	0.92 (0.87, 0.97)	1.12 (1.04, 1.21)	1.02 (0.94, 1.10)	0.85 (0.77, 0.95)	1.00 (0.90, 1.11)
Vanadium					
Odds ratio[b] (95% CI)	0.99 (0.93, 1.04)	0.89 (0.83, 0.95)	0.99 (0.92, 1.07)	0.85 (0.77, 0.94)	1.08 (0.99, 1.19)
Odds ratio[c] (95% CI)	0.91 (0.85, 0.97)	0.92 (0.85, 0.99)	0.96 (0.88, 1.04)	0.88 (0.78, 0.99)	1.13 (1.02, 1.24)
Chromium					
Odds ratio[b] (95% CI)	0.92 (0.88, 0.96)	0.96 (0.91, 1.01)	1.06 (1.00, 1.12)	0.81 (0.74, 0.87)	1.00 (0.93, 1.08)
Odds ratio[c] (95% CI)	0.89 (0.85, 0.93)	0.98 (0.93, 1.03)	1.05 (0.99, 1.11)	0.82 (0.75, 0.89)	1.01 (0.94, 1.09)
Manganese					
Odds ratio[b] (95% CI)	0.95 (0.90, 0.99)	0.98 (0.92, 1.05)	0.99 (0.93, 1.06)	0.86 (0.79, 0.94)	1.16 (1.06, 1.26)
Odds ratio[c] (95% CI)	0.90 (0.86, 0.95)	1.02 (0.95, 1.10)	0.98 (0.91, 1.05)	0.88 (0.80, 0.97)	1.19 (1.08, 1.30)
Iron					
Odds ratio[b] (95% CI)	1.01 (0.95, 1.06)	0.92 (0.85, 0.98)	0.99 (0.92, 1.07)	0.92 (0.84, 1.02)	1.12 (1.02, 1.23)
Odds ratio[c] (95% CI)	0.96 (0.90, 1.02)	0.96 (0.88, 1.04)	0.98 (0.91, 1.06)	0.97 (0.87, 1.08)	1.16 (1.05, 1.28)
Cobalt					
Odds ratio[b] (95% CI)	1.02 (0.97, 1.07)	1.01 (0.95, 1.07)	1.10 (1.03, 1.18)	0.79 (0.72, 0.86)	1.07 (0.98, 1.16)
Odds ratio[c] (95% CI)	0.98 (0.93, 1.04)	1.08 (1.00, 1.15)	1.10 (1.02, 1.18)	0.79 (0.72, 0.88)	1.10 (1.00, 1.20)
Nickel					
Odds ratio[b] (95% CI)	1.07 (1.04, 1.11)	0.94 (0.90, 0.99)	1.04 (1.00, 1.09)	1.00 (0.93, 1.06)	0.92 (0.85, 0.99)
Odds ratio[c] (95% CI)	1.05 (1.01, 1.10)	0.97 (0.92, 1.02)	1.03 (0.98, 1.08)	1.03 (0.96, 1.11)	0.93 (0.86, 1.01)
Copper					
Odds ratio[b] (95% CI)	1.05 (0.97, 1.13)	0.89 (0.81, 0.99)	1.00 (0.90, 1.11)	0.70 (0.61, 0.81)	1.21 (1.07, 1.38)
Odds ratio[c] (95% CI)	1.00 (0.92, 1.09)	0.94 (0.85, 1.04)	0.98 (0.88, 1.10)	0.71 (0.61, 0.83)	1.25 (1.09, 1.44)
Zinc					
Odds ratio[b] (95% CI)	1.07 (0.99, 1.17)	0.91 (0.82, 1.00)	0.82 (0.73, 0.92)	0.80 (0.69, 0.94)	1.29 (1.09, 1.52)
Odds ratio[c] (95% CI)	1.02 (0.93, 1.12)	1.00 (0.89, 1.12)	0.78 (0.69, 0.89)	0.84 (0.70, 1.00)	1.42 (1.18, 1.70)
Arsenic					
Odds ratio[b] (95% CI)	0.97 (0.92, 1.03)	1.05 (0.96, 1.14)	0.98 (0.91, 1.05)	0.92 (0.84, 1.02)	1.05 (0.95, 1.16)
Odds ratio[c] (95% CI)	0.89 (0.83, 0.95)	1.14 (1.03, 1.25)	0.94 (0.86, 1.02)	0.98 (0.88, 1.10)	1.09 (0.98, 1.21)
Selenium					
Odds ratio[b] (95% CI)	1.09 (1.02, 1.17)	0.93 (0.84, 1.03)	1.00 (0.91, 1.11)	0.74 (0.64, 0.86)	1.13 (1.00, 1.26)
Odds ratio[c] (95% CI)	1.06 (0.98, 1.14)	0.96 (0.86, 1.07)	1.00 (0.90, 1.11)	0.76 (0.65, 0.89)	1.15 (1.02, 1.31)
Molybdenum					
Odds ratio[b] (95% CI)	1.08 (1.02, 1.14)	1.00 (0.93, 1.07)	1.05 (0.97, 1.13)	0.86 (0.78, 0.95)	0.91 (0.83, 1.01)
Odds ratio[c] (95% CI)	1.03 (0.97, 1.10)	1.08 (0.99, 1.17)	1.02 (0.93, 1.11)	0.89 (0.79, 0.99)	0.94 (0.84, 1.05)
Cadmium					
Odds ratio[b] (95% CI)	1.02 (0.96, 1.07)	0.96 (0.90, 1.03)	0.98 (0.92, 1.05)	0.85 (0.77, 0.94)	0.98 (0.89, 1.08)
Odds ratio[c] (95% CI)	0.95 (0.89, 1.01)	1.08 (0.99, 1.19)	0.95 (0.88, 1.02)	0.87 (0.77, 0.98)	1.02 (0.92, 1.14)
Thallium					
Odds ratio[b] (95% CI)	1.03 (0.98, 1.08)	1.15 (1.09, 1.21)	1.05 (0.98, 1.12)	0.97 (0.90, 1.06)	0.91 (0.83, 0.99)
Odds ratio[c] (95% CI)	0.97 (0.92, 1.02)	1.35 (1.26, 1.44)	1.01 (0.93, 1.09)	1.04 (0.94, 1.14)	0.94 (0.85, 1.05)
Lead					
Odds ratio[b] (95% CI)	0.99 (0.94, 1.04)	0.94 (0.89, 0.99)	1.06 (0.99, 1.14)	0.73 (0.66, 0.80)	0.98 (0.89, 1.09)
Odds ratio[c] (95% CI)	0.92 (0.86, 0.97)	0.97 (0.91, 1.04)	1.04 (0.95, 1.13)	0.70 (0.62, 0.79)	1.02 (0.91, 1.13)

(Table continues)

Table S2 (continued).

Metal	Colorectal	Breast	Prostate	Stomach	Leukemia
Uranium					
Odds ratio[b] (95% CI)	1.05 (0.99, 1.10)	1.04 (0.97, 1.10)	1.05 (0.97, 1.12)	0.86 (0.78, 0.95)	1.12 (1.01, 1.25)
Odds ratio[c] (95% CI)	1.00 (0.95, 1.07)	1.13 (1.05, 1.22)	1.03 (0.95, 1.11)	0.89 (0.80, 1.00)	1.18 (1.05, 1.31)

Note: CI, confidence interval.
[a]Number of cases and controls from the same provinces (and sex for breast and prostate cancer).
[b]Odds ratio per two-fold increase in measured metal concentrations (95% CI) adjusted for geographical region (12 Spanish provinces), sex (male, female), age (<35 years and 5-year intervals from 35 to 84 years), and educational level (primary or less, high school, and college).
[c]Odds ratio per two-fold increase in measured metal concentrations (95% CI) further adjusted for toenail sample mass (natural cubic spline of log-transformed values with internal knots at the overall 35th and 65th percentiles and boundary knots at the 5th and 95th percentiles).

Table S3. Information criteria and likelihood ratio tests comparing nested models with increasing levels of complexity for measured toenail metal concentrations in a population-based multicase-control study in Spain, 2008–2013 (*n* = 7,923).

	Al	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	As	Se	Mo	Cd	Tl	Pb	U
Akaike information criterion																
Model 1	19,965.9	19,159.7	23,207.8	20,476.3	19,389.6	21,194.3	25,559.6	13,808.8	12,232.5	18,644.9	15,014.9	19,306.6	20,143.2	22,760.7	20,291.0	19,695.6
Model 2	18,078.8	17,105.9	22,514.0	19,525.6	18,053.6	19,781.8	24,542.3	13,016.6	11,048.6	17,032.6	14,444.2	17,265.4	17,597.1	20,024.9	18,338.0	17,940.1
Model 3	18,012.2	17,034.6	22,507.5	19,498.3	18,040.7	19,750.0	24,536.3	12,984.9	11,020.4	16,825.8	14,351.0	17,206.9	17,573.4	20,019.1	18,309.8	17,893.1
Model 4	17,901.6	16,899.3	22,500.8	19,481.2	18,020.5	19,711.3	24,518.7	12,867.2	11,007.5	16,720.5	14,350.5	17,025.7	17,560.7	19,720.3	18,296.4	17,849.6
Model 5	16,822.0	15,428.7	22,290.8	18,847.9	16,896.0	18,568.2	23,660.4	12,003.3	9,560.0	13,817.8	13,185.6	14,524.0	15,595.1	14,586.1	17,716.3	16,710.9
Model 6	16,701.9	15,124.7	22,289.1	18,758.4	16,747.8	18,380.8	23,571.8	11,733.8	9,134.2	13,207.3	12,630.6	13,751.0	15,291.9	13,978.2	17,644.5	16,575.1
Model 7	16,686.3	15,079.0	22,206.6	18,695.8	16,704.9	18,366.1	23,497.1	11,472.5	8,961.0	13,179.6	12,140.7	13,633.6	15,279.2	13,944.0	17,552.2	16,478.6
Bayesian information criterion																
Model 1	20,182.2	19,376.0	23,424.0	20,692.6	19,605.9	21,410.5	25,775.9	14,025.0	12,448.7	18,861.0	15,231.2	19,522.8	20,359.3	22,976.9	20,507.2	19,911.7
Model 2	18,302.1	17,329.1	22,737.3	19,748.9	18,276.8	20,004.9	24,765.6	13,239.8	11,271.9	17,255.7	14,667.5	17,488.6	17,820.2	20,248.1	18,561.2	18,163.2
Model 3	18,249.5	17,271.8	22,744.8	19,735.5	18,277.9	19,987.1	24,773.6	13,222.1	11,257.6	17,062.9	14,588.3	17,444.1	17,810.5	20,256.3	18,547.0	18,130.1
Model 4	18,159.8	17,157.5	22,759.0	19,739.3	18,278.6	19,969.3	24,776.8	13,125.3	11,265.6	16,978.5	14,608.7	17,283.8	17,818.7	19,978.4	18,554.5	18,107.5
Model 5	17,087.1	15,693.8	22,556.0	19,113.0	17,161.1	18,833.2	23,925.5	12,268.3	9,825.1	14,082.8	13,450.7	14,789.1	15,860.0	14,851.1	17,981.4	16,975.8
Model 6	17,029.8	15,452.6	22,617.1	19,086.3	17,075.7	18,708.6	23,899.7	12,061.7	9,462.1	13,535.0	12,958.5	14,078.9	15,619.6	14,306.0	17,972.3	16,902.7
Model 7	17,035.2	15,427.9	22,555.5	19,044.7	17,053.7	18,714.7	23,846.0	11,821.2	9,309.8	13,528.2	12,489.5	13,982.4	15,627.8	14,292.7	17,900.9	16,827.1
Likelihood ratio test *p*-value																
Model 2 vs 1	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Model 3 vs 2	<0.001	<0.001	0.01	<0.001	<0.001	<0.001	0.01	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.01	<0.001	<0.001
Model 4 vs 3	<0.001	<0.001	0.01	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.09	<0.001	<0.001	<0.001	<0.001	<0.001
Model 5 vs 4	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Model 6 vs 5	<0.001	<0.001	0.02	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Model 7 vs 6	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

Note: All models were fitted using maximum likelihood methods to allow comparison of models with different fixed-effects specifications. Model 1 only included fixed effects for case-control status and sociodemographic factors, model 2 extended model 1 with a fixed linear term for individual log-transformed toenail sample mass, model 3 extended model 1 with fixed natural cubic splines for individual log-transformed mass, and model 4 extended model 3 with fixed natural cubic splines for mean log-transformed mass of each laboratory batch. Among mixed models, model 5 extended model 4 with random intercepts across laboratory batches and model 6 extended model 5 with random between-batch natural cubic splines for individual log-transformed mass. Finally, model 7 generalized model 6 to heterogeneous within-batch error variance as natural cubic splines of log-transformed mass. Al, aluminum; As, arsenic; Cd, cadmium; Co, cobalt; Cr, chromium; Cu, copper; Fe, iron; Mn, manganese; Mo, molybdenum; Ni, nickel; Pb, lead; Se, selenium; Tl, thallium; U, uranium; V, vanadium; Zn, zinc.

Table S4. Odds ratios for common cancers per two-fold increase in calibrated toenail metal concentrations in a population-based multicase-control study in Spain, 2008–2013 (*n* = 7,923).

Metal	**Colorectal**	**Breast**	**Prostate**	**Stomach**	**Leukemia**
No. of cases/controls[a]	1,422/3,373	1,438/1,679	915/1,273	343/2,728	355/1,524
Aluminum (ppm)					
Geometric mean[b] (SD)					
Cases	36.4 (2.1)	51.7 (1.7)	31.5 (2.0)	36.4 (2.1)	35.3 (2.2)
Controls	39.7 (2.0)	50.8 (1.8)	28.9 (2.1)	38.5 (2.0)	35.6 (2.1)
Odds ratio[c] (95% CI)	0.90 (0.85, 0.97)	0.99 (0.90, 1.08)	1.12 (1.03, 1.22)	0.95 (0.84, 1.07)	1.04 (0.92, 1.17)
Vanadium (ppb)					
Geometric mean[b] (SD)					
Cases	54.7 (1.9)	69.3 (1.6)	51.7 (2.0)	52.6 (1.8)	65.3 (2.4)
Controls	58.7 (1.9)	68.9 (1.7)	48.0 (1.9)	57.1 (1.9)	59.9 (1.8)
Odds ratio[c] (95% CI)	0.89 (0.82, 0.96)	0.94 (0.85, 1.05)	1.13 (1.02, 1.24)	0.90 (0.78, 1.03)	1.14 (1.01, 1.28)
Chromium (ppm)					
Geometric mean[b] (SD)					
Cases	1.08 (2.8)	1.85 (2.4)	1.06 (2.6)	0.91 (2.7)	1.45 (3.7)
Controls	1.45 (2.8)	2.05 (2.5)	1.00 (2.7)	1.38 (2.9)	1.44 (2.7)
Odds ratio[c] (95% CI)	0.87 (0.83, 0.92)	0.93 (0.87, 0.98)	1.05 (0.99, 1.12)	0.83 (0.76, 0.91)	1.03 (0.95, 1.11)
Manganese (ppb)					
Geometric mean[b] (SD)					
Cases	484 (2.4)	559 (1.9)	479 (2.4)	444 (2.2)	526 (3.0)
Controls	501 (2.3)	555 (2.0)	419 (2.4)	487 (2.4)	501 (2.3)
Odds ratio[c] (95% CI)	0.91 (0.86, 0.96)	0.96 (0.88, 1.03)	1.11 (1.03, 1.20)	0.87 (0.79, 0.97)	1.17 (1.06, 1.29)
Iron (ppm)					
Geometric mean[b] (SD)					
Cases	28.7 (2.1)	37.8 (1.7)	25.9 (2.1)	27.0 (1.9)	32.2 (2.6)
Controls	29.9 (2.0)	36.5 (1.8)	23.3 (2.1)	28.7 (2.0)	29.7 (2.0)
Odds ratio[c] (95% CI)	0.95 (0.89, 1.02)	1.02 (0.93, 1.13)	1.15 (1.06, 1.25)	0.95 (0.85, 1.08)	1.14 (1.02, 1.27)
Cobalt (ppb)					
Geometric mean[b] (SD)					
Cases	24.0 (2.2)	31.0 (1.9)	19.2 (2.4)	20.5 (2.2)	26.6 (2.7)
Controls	24.1 (2.2)	30.4 (1.9)	18.1 (2.3)	22.9 (2.2)	24.8 (2.1)
Odds ratio[c] (95% CI)	1.01 (0.95, 1.07)	1.00 (0.92, 1.09)	1.08 (1.00, 1.17)	0.92 (0.82, 1.02)	1.13 (1.02, 1.25)
Nickel (ppm)					
Geometric mean[b] (SD)					
Cases	2.02 (3.3)	2.85 (2.4)	1.65 (3.1)	1.99 (3.3)	1.72 (3.1)
Controls	1.98 (2.9)	2.90 (2.4)	1.44 (3.0)	2.11 (3.0)	1.91 (2.9)
Odds ratio[c] (95% CI)	1.08 (1.03, 1.12)	0.99 (0.93, 1.05)	1.06 (1.00, 1.12)	1.04 (0.97, 1.12)	0.98 (0.90, 1.07)
Copper (ppm)					
Geometric mean[b] (SD)					
Cases	4.25 (1.6)	4.86 (1.4)	4.48 (1.6)	3.95 (1.5)	4.81 (1.8)
Controls	4.50 (1.6)	4.78 (1.5)	4.22 (1.6)	4.49 (1.6)	4.29 (1.5)
Odds ratio[c] (95% CI)	0.93 (0.83, 1.03)	0.97 (0.84, 1.12)	1.19 (1.04, 1.36)	0.66 (0.55, 0.80)	1.43 (1.19, 1.72)
Zinc (ppm)					
Geometric mean[b] (SD)					
Cases	102 (1.4)	100 (1.4)	108 (1.5)	97 (1.4)	107 (1.6)
Controls	102 (1.5)	100 (1.4)	105 (1.5)	103 (1.4)	102 (1.4)
Odds ratio[c] (95% CI)	1.06 (0.94, 1.19)	1.02 (0.87, 1.19)	1.14 (0.99, 1.33)	0.78 (0.63, 0.98)	1.09 (0.89, 1.34)

(Table continues)

Table S4 (continued).

Metal	Colorectal	Breast	Prostate	Stomach	Leukemia
Arsenic (ppb)					
Geometric mean[b] (SD)					
Cases	196 (1.8)	153 (1.6)	253 (1.7)	195 (1.8)	209 (2.1)
Controls	184 (1.7)	149 (1.5)	234 (1.7)	182 (1.7)	199 (1.8)
Odds ratio[c] (95% CI)	1.03 (0.94, 1.12)	1.18 (1.05, 1.34)	1.26 (1.12, 1.41)	1.08 (0.92, 1.28)	1.02 (0.89, 1.17)
Selenium (ppm)					
Geometric mean[b] (SD)					
Cases	0.79 (1.7)	0.79 (1.4)	0.86 (1.6)	0.72 (1.6)	0.82 (1.9)
Controls	0.79 (1.5)	0.78 (1.3)	0.82 (1.7)	0.79 (1.5)	0.78 (1.6)
Odds ratio[c] (95% CI)	1.13 (1.02, 1.24)	1.18 (1.01, 1.40)	1.16 (1.02, 1.32)	0.81 (0.67, 0.99)	1.12 (0.95, 1.31)
Molybdenum (ppb)					
Geometric mean[b] (SD)					
Cases	41.5 (1.9)	51.0 (1.6)	34.7 (1.8)	36.8 (1.9)	35.9 (2.0)
Controls	41.5 (1.8)	50.3 (1.6)	33.2 (1.8)	39.7 (1.8)	39.1 (1.8)
Odds ratio[c] (95% CI)	1.05 (0.97, 1.14)	1.03 (0.91, 1.15)	1.17 (1.05, 1.31)	0.97 (0.84, 1.12)	0.97 (0.83, 1.14)
Cadmium (ppb)					
Geometric mean[b] (SD)					
Cases	15.4 (2.0)	19.2 (1.5)	13.5 (2.0)	14.5 (1.8)	15.7 (2.2)
Controls	15.9 (1.8)	18.7 (1.5)	13.1 (2.0)	15.5 (1.8)	14.9 (1.8)
Odds ratio[c] (95% CI)	1.02 (0.95, 1.09)	1.05 (0.93, 1.19)	1.05 (0.96, 1.15)	0.97 (0.84, 1.11)	1.12 (0.99, 1.27)
Thallium (ppb)					
Geometric mean[b] (SD)					
Cases	4.14 (1.8)	4.14 (1.7)	4.25 (1.8)	3.95 (1.6)	4.59 (1.8)
Controls	4.01 (1.7)	3.95 (1.6)	4.12 (1.7)	3.98 (1.7)	4.19 (1.7)
Odds ratio[c] (95% CI)	1.05 (0.97, 1.15)	1.13 (1.02, 1.26)	1.09 (0.97, 1.22)	0.97 (0.83, 1.14)	1.14 (0.98, 1.32)
Lead (ppb)					
Geometric mean[b] (SD)					
Cases	616 (1.9)	681 (2.0)	644 (2.0)	554 (2.0)	697 (2.1)
Controls	655 (2.0)	677 (2.0)	607 (1.9)	643 (2.0)	671 (1.9)
Odds ratio[c] (95% CI)	0.88 (0.82, 0.94)	0.97 (0.90, 1.04)	1.08 (0.99, 1.19)	0.77 (0.68, 0.88)	1.03 (0.90, 1.16)
Uranium (ppb)					
Geometric mean[b] (SD)					
Cases	6.44 (2.0)	6.95 (1.9)	7.66 (2.1)	5.77 (1.9)	9.66 (2.1)
Controls	6.47 (2.0)	6.50 (1.8)	6.91 (2.1)	6.65 (1.9)	8.14 (1.9)
Odds ratio[c] (95% CI)	1.07 (0.99, 1.15)	1.19 (1.08, 1.30)	1.20 (1.10, 1.32)	0.91 (0.79, 1.04)	1.15 (1.01, 1.32)

Note: CI, confidence interval; ppb, parts per billion; ppm, parts per million; SD, standard deviation.
[a]Number of cases and controls from the same provinces (and sex for breast and prostate cancer).
[b]Geometric mean (geometric SD) of calibrated metal concentrations among cases and controls.
[c]Odds ratio per two-fold increase in calibrated metal concentrations (95% CI) adjusted for geographical region (12 Spanish provinces), sex (male, female), age (<35 years and 5-year intervals from 35 to 84 years), and educational level (primary or less, high school, and college).

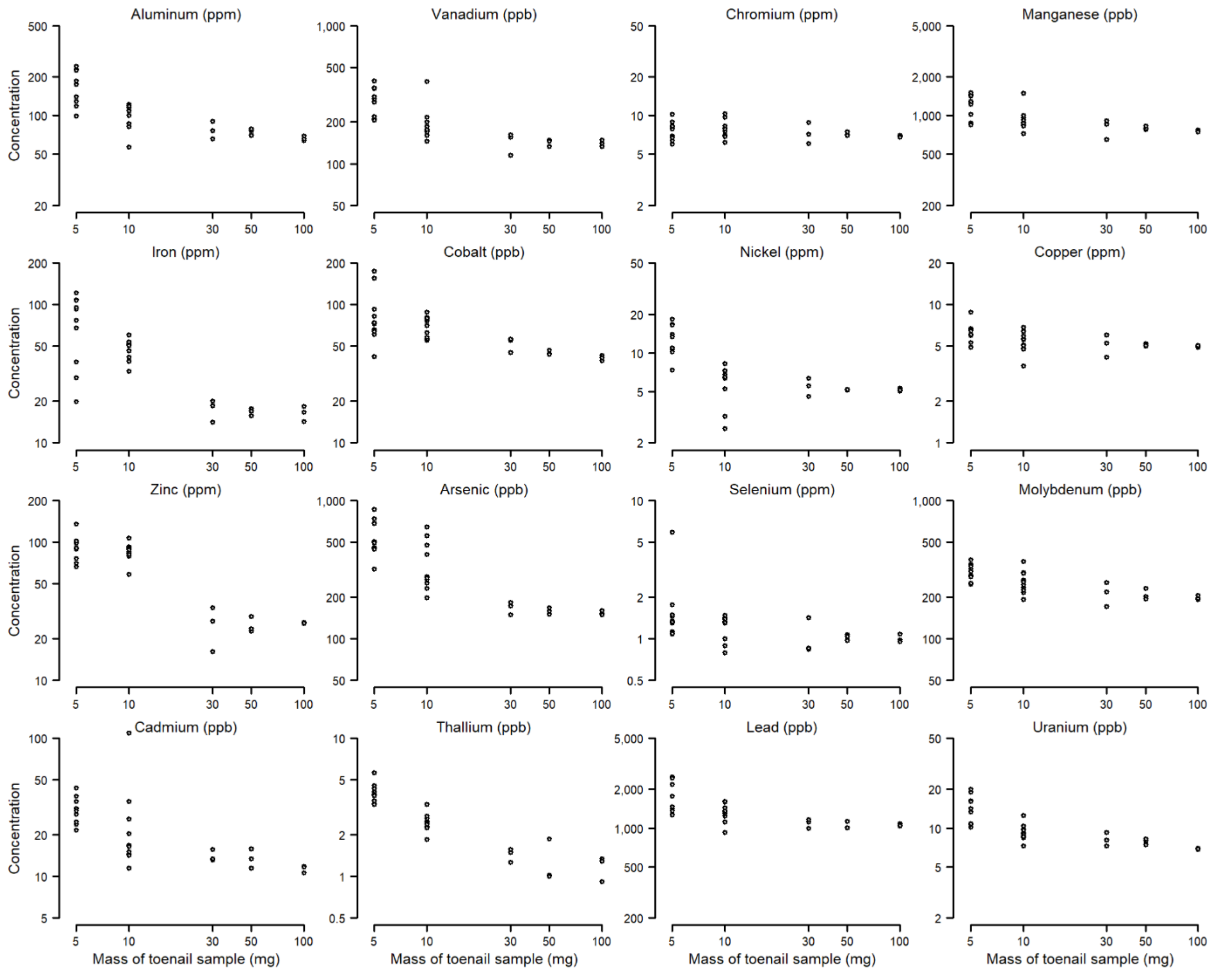

Figure S1. Measured metal concentrations in homogenized toenail samples of different mass in the reproducibility study (*n* = 29). Results were based on a pooled sample of toenails, which was cryohomogenized and the resulting powder divided into ten samples of 5 and 10 mg and three samples of 30, 50, and 100 mg. Note: ppb, parts per billion; ppm, parts per million.

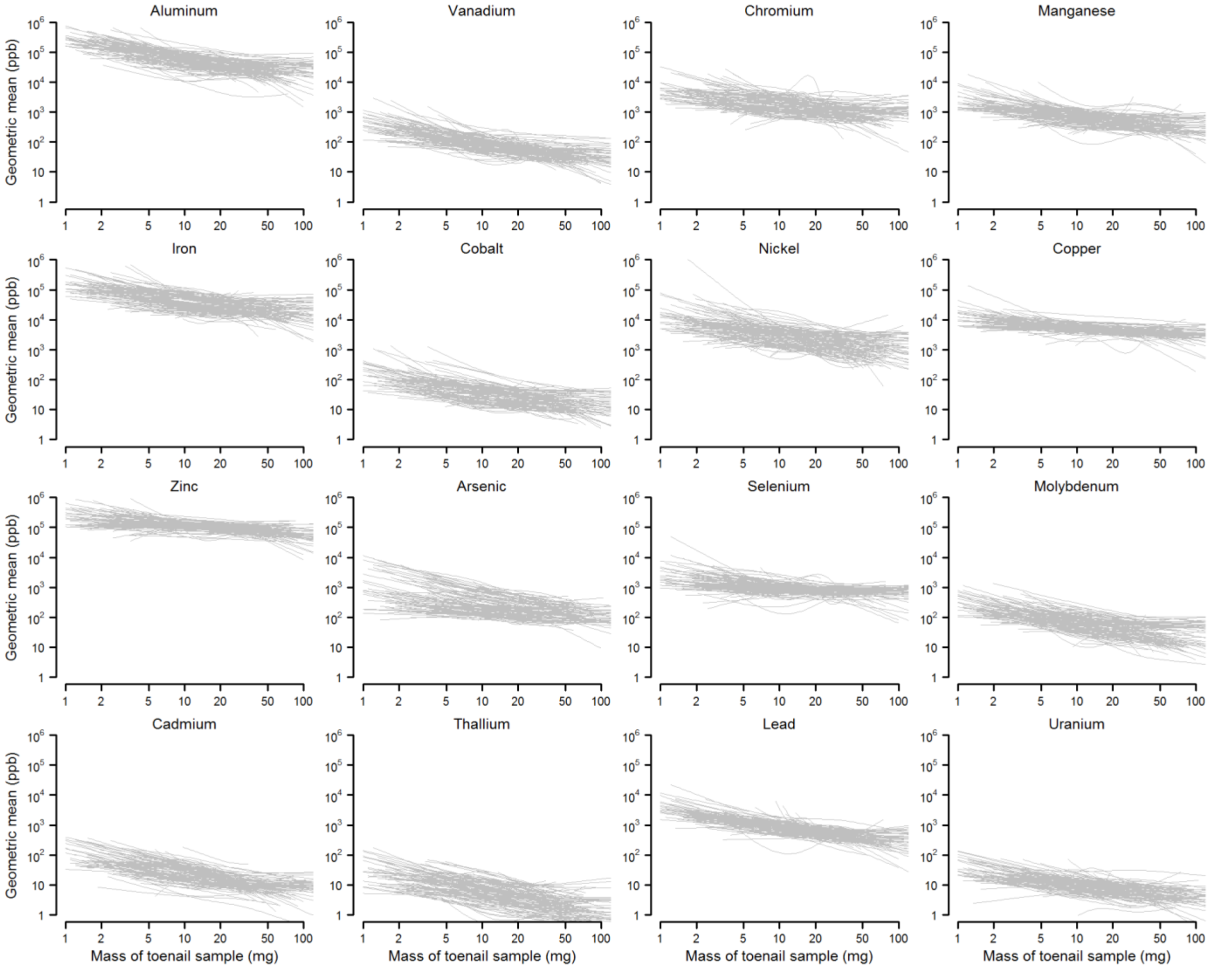

Figure S2. Geometric mean metal concentrations as a smooth function of toenail sample mass within each laboratory batch in a population-based multicase-control study in Spain, 2008–2013 (n = 7,923). Batch-specific trends in geometric mean metal concentrations (gray curves) were obtained from separate regression models within each laboratory batch relating log-transformed metal concentrations with a natural cubic spline of log-transformed toenail sample mass with internal knots at the overall 35th and 65th percentiles and boundary knots at the 5th and 95th percentiles (3.0, 13.9, 28.9, and 63.2 mg, respectively). Batch-specific analyses were restricted to 119 of the 122 laboratory batches (97.5%) with at least 25 metal determinations. Note: ppb, parts per billion.

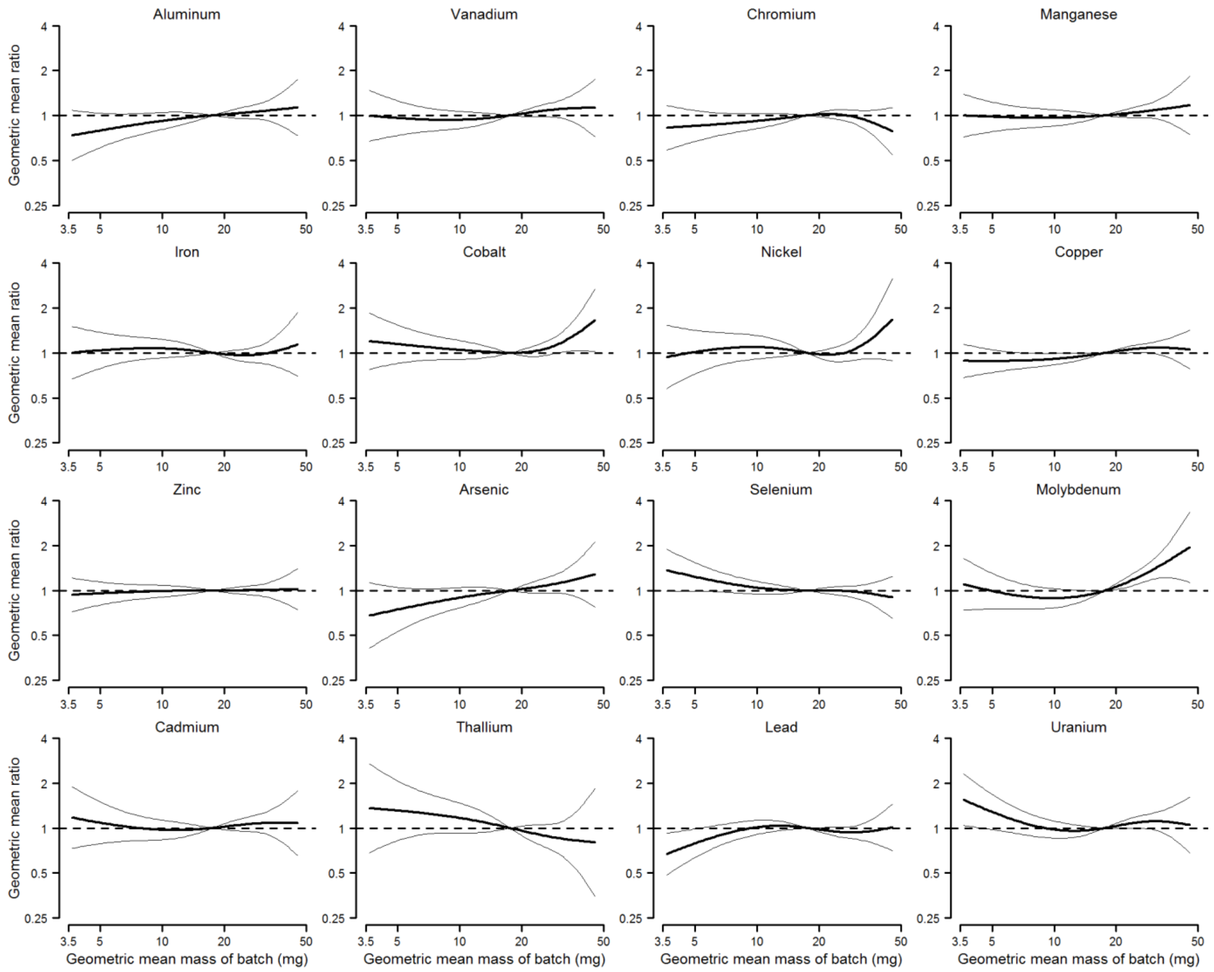

Figure S3. Batch compositional bias in measured metal concentrations associated with toenail samples of the same mass analyzed in laboratory batches with different geometric mean masses in a population-based multicase-control study in Spain, 2008–2013 (*n* = 7,923). The batch compositional bias in measured metal concentrations (thick curve) and its 95% confidence interval (thin curves) were determined by the estimates and standard errors of the fixed spline parameters η_1, η_2, and η_3 associated with mean log-transformed masses for each laboratory batch in heteroscedastic spline mixed models for log-transformed metal concentrations. The reference level was set at the geometric mean mass over all laboratory batches (17.6 mg).

References

1. Raudenbush SW, Bryk AS. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*. 2nd ed. Thousand Oaks, CA: Sage Publications.
2. Pinheiro JC, Bates DM. 2000. Extending the basic linear mixed-effects model. In: *Mixed-Effects Models in S and S-PLUS*. Pinheiro JC, Bates DM, eds. New York, NY: Springer, 201–270.
3. Durrleman S, Simon R. 1989. Flexible regression models with cubic splines. Stat Med 8(5):551–561, PMID: 2657958, https://doi.org/10.1002/sim.4780080504.
4. Searle SR. 1982. *Matrix Algebra Useful for Statistics*. New York, NY: John Wiley & Sons.
5. Snijders TAB, Bosker RJ. 2012. How much does the model explain? In: *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. Snijders TAB, Bosker RJ, eds. 2nd ed. London: Sage Publications, 109–118.