Annals of Work Exposures and Health, 2024, **XX**, 1–4
https://doi.org/10.1093/annweh/wxae065
Advance access publication 16 August 2024
Correction

The Chartered Society for Worker Health Protection
OXFORD

Correction to: Bayesian Analysis of Occupational Exposure Data with Conjugate Priors

This is a correction to: Rachael M. Jones, Igor Burstyn, Bayesian Analysis of Occupational Exposure Data with Conjugate Priors, Annals of Work Exposures and Health, Volume 61, Issue 5, June 2017, Pages 504–514, https://doi.org/10.1093/annweh/wxx032

This correction addresses an error in the equations and computer code related to Method 2 presentation in the original article. The below details have been corrected only in this correction notice to preserve the published version of record.

Method II

We present a second Bayesian analysis method that involves a normal and an inverse-Γ conjugate prior distribution: With a normal likelihood function, this prior distribution gives a joint posterior distribution with the same form as the prior distribution (normal-inverse-Γ) (Murphy 2007). Using notation previously defined, the joint prior distribution for μ and σ^2 is Normal-Inverse-Γ($\mu_0, \kappa_0, a_0, b_0$), which can be decomposed into:

Eq: 7 $\sigma^2 \sim \text{Inv-}\Gamma(a_0, b_0)$
Eq: 8 $\mu \sim N(\mu_0, \sigma^2/\kappa_0)$

where $a_0 = \kappa_0/2$ and $b_0 = \kappa_0\sigma_0^2/2$ represent the inverse-Γ shape and rate parameters, respectively. The joint posterior distribution is obtained by multiplying the prior distribution by normal likelihood. The posterior marginal distributions is Normal-Inverse-Γ($\mu_n, \kappa_n, a_n, b_n$), where $\mu_n=(\kappa_0\mu_0 + n\bar{y})/ (\kappa_0 + n)$, $\kappa_n=\kappa_0 + n$, $a_n = (\kappa_0 + n)/2$, and $b_n=0.5(\kappa_0\sigma_0^2 + nS_y^2 + (\kappa_0 n(\mu_0- \bar{y})^2)/(\kappa_0 + n))$ (the rate parameter). The features of the posterior mean and variance distributions can be calculated by drawing samples from the posterior distribution, which is operationalized in two steps, using easily accessible software, such as a spreadsheet. These steps are first to draw variance σ_y^2 from Inverse-Γ(a_n, b_n), followed by a draw of μ_y from N(μ_n, σ_y^2/κ_n).

We updated all code in the supplementary materials, which we reproduce in their entirety for convenience (Supplementary Material A).

An example

In applying Method II, data from 1985 were used to define the prior distribution and pooled data from 1987 and 1989 were used to define the likelihood. The parameter values calculated from the data for the Bayesian analysis included those described previously, and method-specific parameters in the posterior distribution were calculated: μ_n = -2.13, κ_n = 80, a_n = 40, and b_n = 31.2. One hundred thousand values (B = 100,000) were drawn from the posterior, and (after transformation from the log-scale) are shown in Figure 2 and summarized in Table III. The mean values of the posterior distributions in Method II are similar to those from Method I. With the correction, the 95%CrI are similar for the two methods. Applying the exposure categories (Table I) to the posterior distribution values of $X_{0.95}$ (Table III) resulted in the following multinomial distribution: P(Category 0) = P(Category 1) = P(Category 2) = P(Category 3) = 0 and P(Category 4) = 1.

Comment II: Towards a default prior on variance

We consider the inverse-Γ distribution fitted to the percentiles reported for between-worker variance when workers were classified across jobs by location, inverse-Γ(a = 6.25, b = 4.80). This worker classification was selected because our example data involve single exposure measurements for workers with unspecified jobs working in a single foundry. This default inverse-Γ prior distribution has a larger variance (0.20 vs. 0.07) and smaller mean (0.91 vs. 1.51) than the prior based on data, inverse-Γ(a_0 = 35, b_0 = 51.5). Results for the Method II analysis using the default prior for variance (e.g., a_0 = 6.25, b_0 = 4.80) and the mean and sample size consistent with the prior data (μ_0 = 0.20, κ_0 =70) are shown in Table III. As expected from the features of the default and data-based prior distributions, the posterior distribution for variance has a larger mean and variance, which results in higher values for the posterior mean distribution of lead exposure and the $X_{0.95}$. Applying the exposure categories (Table 1) to the posterior distribution values of $X_{0.95}$ (Table 3) resulted in the following multinomial distribution: P(Category 0) = P(Category 1) = P(Category 2) = P(Category 3) = 0 and P(Category 4) = 1.

Acknowledgments

The authors thank Dr. Harrison Quick for “statistical first aid”.

References

Murphy KP. 2007. Conjugate Bayesian analysis of the Gaussian distribution. [accessed 2016 Nov 8]. https://www.cs.ubc.ca/~murphyk/Papers/bayesGauss.pdf.

Table 3. Mean and 95% credible intervals (95%CrI) of the posterior distributions of key parameters in the distribution of lead exposure data (μg/L) using an informative conjugate priors (Method I and Method II) and a vague prior distribution.

Exposure distribution parameter	Posterior distribution							
	Method I		Method II		Method III		Method II with default prior	
	Mean	95%CrI	Mean	95%CrI	Mean	95%CrI	Mean	95%CI
Mean (μg/L)	0.20	0.15, 0.26	0.18	0.14, 0.23	0.31	0.04, 0.50	0.48	0.22, 1.25
GM (μg/L)	0.12	0.10, 0.15	0.12	0.097, 0.14	0.04	0.01, 0.08	0.12	0.08, 0.17
GSD	2.74	2.36, 3.24	2.44	2.15, 2.84	4.06	2.29, 8.94	4.96	3.23, 8.48
95th%tile (μg/L)	0.63	0.45, 0.89	0.52	0.39, 0.71	0.47	0.10, 1.71	1.75	0.76, 4.33

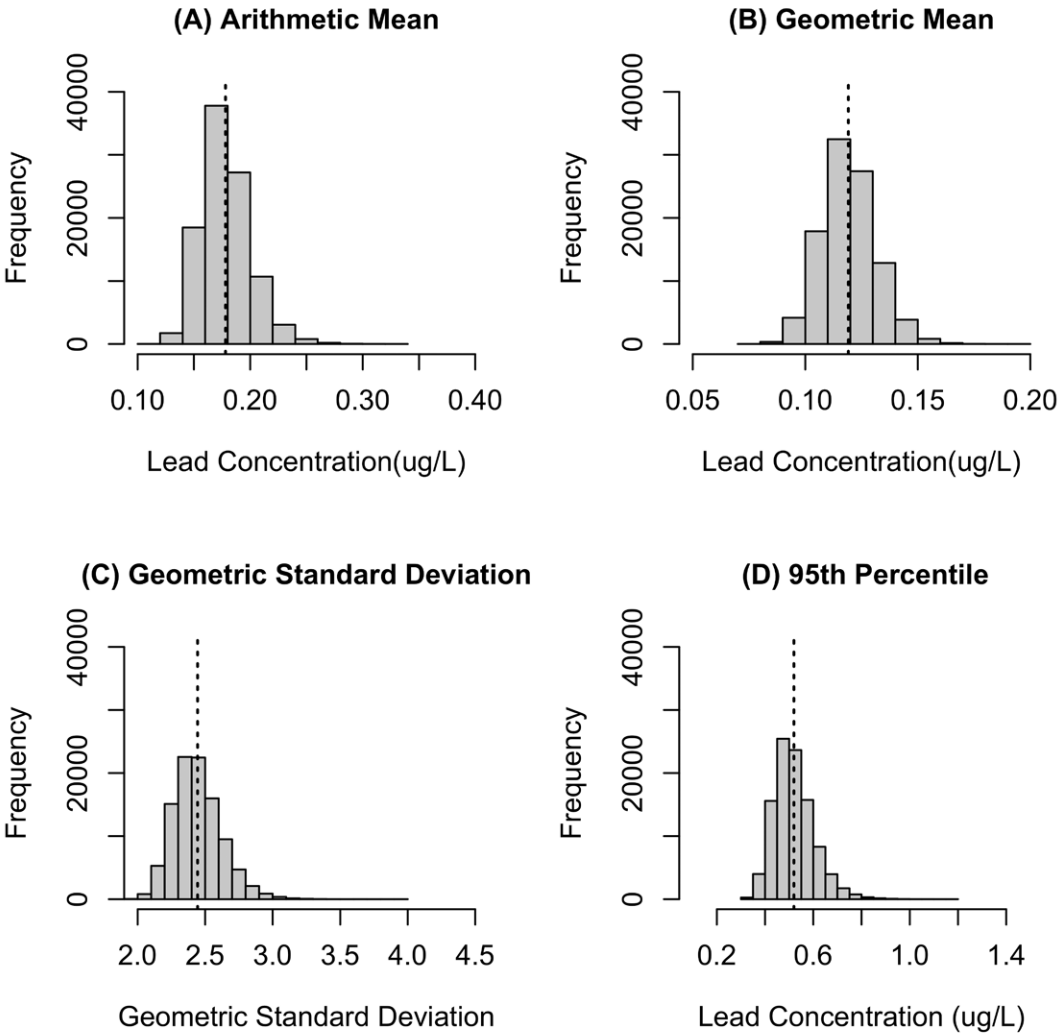

Fig. 2. Posterior distributions of key parameters of the lead exposure distributions using Method II. Vertical dotted lines indicate the mean values.

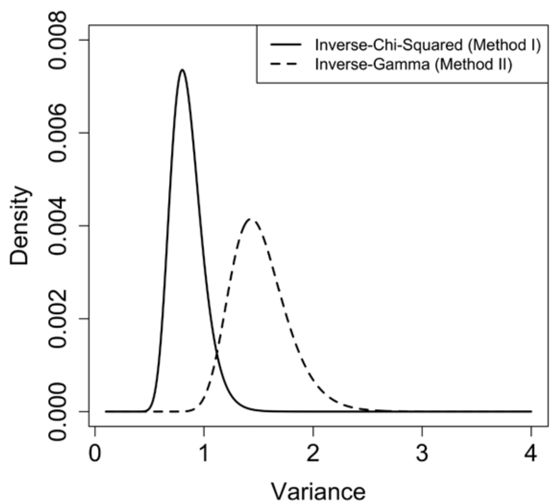

Fig. 3. Prior distributions for the variance parameter in the example.

Supplemental Materials A

This function was written in the R Project for Statistical Computing to implement the normal- inverse-Γ conjugate prior distribution for normal data described by Murphy (2007) (Method II). The function, named M2, takes four values as inputs: p.data is a vector of exposure data to define the prior distribution, l.data is a vector of exposure data to define the likelihood, B is a number defining how many draws from the posterior distribution that will be made, and OEL is a number defining the occupational exposure limit and is an optional parameter. The output of the function is a data frame of size B rows and 6 or 7 columns, containing draws from the posterior distributions of σ_y^2/k_n (sigma2p) and μ_y (mup), and calculated values of the mean (meanp), GM (GMp), GSD (GSDp), $X_{0.95}$ (X95p), and the exposure category (Ecat). Ecat is only included if OEL was specified.

```
M2<-function(p.data, l.data, B, OEL){

     #set up the prior components
     kappa0<-length(p.data)



     tau0<-1/var(log(p.data))
     mu0<-mean(log(p.data))
     a0<- kappa0/2
     b0<- kappa0/2*tau0

     #set up the likelihood components from data
     n<-length(l.data)
     ybar<-mean(log(l.data))
     Sy2<-var(log(l.data)

     #parameters of the posterior

     mun<-(kappa0*mu0+n*ybar)/(kappa0+n)
     kn=kappa0+n
     an<-a0+0.5*n
     bn=0.5*(kappa0/tau0 + n*Sy2 + (kappa0*n*(mu0-mun)^2)/(kappa0+n))

     #sample from the posterior


      sigma2y=rgamma(B, shape=an, rate=bn)
        sigma2p=sigma2y/kn
muy=rnorm(B, mun, sqrt(sigma2p))

     #calculate other posterior statistics
     GMp=exp(muy)
     GSDp=exp(sqrt(sigma2p))
     meanp<-GMp*exp(0.5*sigma2p)
     X95p=GMp*GSDp^1.645

     if(missing(OEL)){
         pparam<-data.frame(GMp, GSDp, X95p, sigma2p, muy, meanp)
     } else {
         Ecat<-ifelse(X95p>OEL,4,
         ifelse(X95p<=OEL & X95p>0.5*OEL,3,
         ifelse(X95p<=0.5*OEL & X95p>0.1*OEL, 2,
         ifelse (X95p<=0.1*OEL & X95p>0.01*OEL,1, 0))))
         pparam<-data.frame(GMp, GSDp, X95p, sigma2p, muy, meanp, Ecat)
     }

     #Calculate predicted proportion of exposures in each exposure
category
     colnames(pparam)=c("GM", "GSD", "X95", "sigma2p", "muy", "AM",
"exposure_category")

      table<-prop.table(table(pparam$Ecat))
      return(list(pparam, table))}
```