

## SUPPLEMENTARY MATERIAL

### Binary Regression Models with Log-Link in the Cohort Studies

Katri Jalava<sup>1</sup>, Sirpa Räsänen<sup>2</sup>, Kaija Ala-Kojola<sup>2</sup>, Saara Nironen<sup>3</sup>, Jyrki Möttönen<sup>4</sup> and Jukka Ollgren<sup>1</sup>

<sup>1</sup>Department of Infectious Disease Surveillance and Control, National Institute for Health and Welfare, Helsinki, Finland

<sup>2</sup>Health services, City of Tampere, Tampere, Finland

<sup>3</sup>Health services, Rauma Town, Rauma, Finland

<sup>4</sup>Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland

#### Supplementary Material 1. Outbreak 1 Data for the Quantity of the Pesto Chicken as an Excel Spreadsheet

Case 1	pestochickenq
1	3.00
1	2.00
1	3.00
1	3.00
1	1.00
0	1.00
1	1.00
1	3.00
1	3.00
1	3.00
1	1.00
1	1.00
1	1.00
1	3.00
1	1.00
1	3.00
1	3.00
1	3.00
1	3.00
1	1.00
1	1.00
1	1.00
1	3.00
1	3.00
1	3.00
1	2.00
1	1.00
1	3.00
1	3.00
1	1.00
1	3.00
1	3.00
1	3.00

1	3.00
1	1.00
1	3.00
1	2.00
0	2.00
0	3.00
0	1.00
0	1.00
0	3.00
0	1.00
0	1.00
0	1.00
0	1.00
0	1.00
0	1.00
0	1.00
0	3.00
0	3.00
0	1.00
0	2.00
0	3.00
0	1.00
0	3.00

**Supplementary Material 2. Log-Binomial Model for the Outbreak 1 as a Text File for Winbugs**

```

model {

  for (i in 1:N){
    D.pc2[i]<-equals(pestochickenq[i],2)
    D.pc3[i]<-equals(pestochickenq[i],3)

    for (i in 1:N){
      case1[i]~dbern(p[i])

      log(p[i])<-alpha+beta1*D.pc2[i]+beta2*D.pc3[i]
    }

    alpha~dnorm(0,0.001)
    beta1~dnorm(0,0.001)
    beta2~dnorm(0,0.001)

    RR_pestochicken1<-exp(beta1)
    RR_pestochicken2<-exp(beta2)

  }

```

**Supplementary Material 3. SPSS Syntax for the R to Winbugs Connection for the Outbreak 1**

```

BEGIN PROGRAM R.
library(R2WinBUGS)

N=61

casedata <- spssdata.GetDataFromSPSS()
attach(casedata)

data=list("N","case1","pestochickenq")
parameters<-c("alpha","RR_pestochicken1","RR_pestochicken2")
inits=function() {list(alpha= 0.0, beta1=0.0, beta2=0.0)}

model.sim<-bugs(data, inits, parameters, "modell1.bug",
  n.chains=1, n.iter=50000, n.thin=1, working.directory="c:/",
  debug=TRUE, bugs.directory="c:/Program Files/WinBUGS14/")
print(model.sim)

spssdata.CloseDataConnection()

END PROGRAM.

```

**Supplementary Material 4. Outbreak 2 Data for the Cohort as an Excel Spreadsheet**

Case	Fish	Pasta
0	0	0
1	1	1
1	1	1
1	1	0
1	1	1
0	1	1
1	1	0
1	1	1
0	0	0
0	1	0
1	1	1
1	1	1
1	1	1
1	1	1
1	1	1
1	1	1
1	1	1
1	1	1
1	1	0
0	1	1

1	1	1
1	1	0
1	1	1
1	1	1
0	1	1
0	0	0
0	1	1
1	1	1
0	0	0
0	0	0
1	0	0
0	0	0
1	1	1
0	0	1
0	0	0
0	0	0
1	0	1
1	1	0
1	1	1
0	0	0
0	0	0
1	1	0
1	0	0
0	0	0
0	1	0
0	0	0
1	0	1
1	0	0
1	0	1
0	0	0
0	0	0
0	0	0
1	0	0
0	0	0
0	0	0
1	0	0
1	0	0

**Supplementary Material 5. Log-Binomial Model for the Outbreak 2 as a Text File for Winbugs**

```
model {  
  
  for (i in 1:N){  
    case[i]~dbern(p[i])  
  
    log(p[i])<-alpha+beta1*fish[i]+ beta2*pasta[i]  
  
  }  
  
  alpha~dnorm(0,0.001)  
  beta1~dnorm(0,0.001)  
  beta2~dnorm(0,0.001)  
  
  RR_fish<-exp(beta1)  
  RR_pasta<-exp(beta2)  
  
}
```

**Supplementary Material 6. SPSS Syntax for the R to Winbugs Connection for the Outbreak 2**

```
BEGIN PROGRAM R.  
library(R2WinBUGS)  
  
N=70  
  
casedata <- spssdata.GetDataFromSPSS()  
attach(casedata)  
  
data=list("N","case","fish","pasta")  
parameters<-c("alpha","RR_fish", "RR_pasta")  
inits=function() {list(alpha= 0.0, beta1=0.0, beta2=0.0)}  
  
model.sim<-bugs(data, inits, parameters, "model2.bug",  
  n.chains=1, n.iter=25000, n.thin=1, working.directory="c:/",  
debug=FALSE, bugs.directory="c:/Program Files/WinBUGS14/")  
print(model.sim)  
  
spssdata.CloseDataConnection()  
  
END PROGRAM.
```

**Supplementary material 7. Mathematical Proof for the Invalidity of the Conversion Formula as Suggested Previously**

$$\begin{aligned}
RR &= \frac{p1}{p2} = \frac{\frac{e^{\mu+\beta_1*(x_1+1)+\beta_2*x_2}}{1 + e^{\mu+\beta_1*(x_1+1)+\beta_2*x_2}}}{\frac{e^{\mu+\beta_1*x_1+\beta_2*x_2}}{1 + e^{\mu+\beta_1*x_1+\beta_2*x_2}}} = \frac{e^{\beta_1} * (1 + e^{\mu+\beta_1*x_1+\beta_2*x_2})}{1 + e^{\mu+\beta_1*(x_1+1)+\beta_2*x_2}} \\
&= \frac{e^{\beta_1} * (1 + e^{\mu+\beta_1*x_1+\beta_2*x_2})}{1 + e^{\beta_1}e^{\mu+\beta_1*x_1+\beta_2*x_2}} \quad (1) \\
&= \frac{e^{\beta_1} \left( \frac{1}{1-p} \right)}{\frac{1}{1-p} (1-p + pe^{\beta_1})} = \frac{e^{\beta_1}}{(1-p) + pe^{\beta_1}} \stackrel{(2)}{=} \frac{OR}{(1-p) + p * OR}
\end{aligned}$$

if  $p \approx 0 \implies RR = OR$

Note:

$$e^{\mu+\beta_1*x_1+\beta_2*x_2} = \frac{p}{1-p} \quad (1) \quad e^{\beta_1} = OR \quad (2)$$

As it was assumed above in (1) that  $p$  depends on all covariates  $x$ , the final equation as originally suggested by Zhang et al., is not universally valid.