7th International Conference on Predictive Modelling of Food Quality and Safety

7ICPMF, September 12 — 15th, 2011, Dublin, Ireland

Radisson Blu Royal Hotel, Golden Lane, Dublin 8, Ireland

Conference Proceedings

www.icpmf.org/2011

E. Cummins, J.M. Frías and V.P. Valdramidis (Eds.),
Predictive Modelling of Food Quality and Safety – Conference Proceedings,
UCD, DIT, Teagasc, Dublin, Ireland

High number of servings reduces the coefficient of variation of food-borne burden-of-illness

F. Pérez-Rodríguez[1] and M.H. Zwietering[2]

[1] Department of Food Science and Technology, University of Córdoba, Campus de Rabanales, C-1, 14014 Córdoba, Spain
[2] Laboratory of Food Microbiology, Wageningen University, P.O. Box 8129, 6700 EV Wageningen, The Netherlands

Abstract

The Central Limit Theorem (CLT) is proposed as a means of understanding microbial risk in foods from a Public Health perspective. On the basis of the CLT, the hypothesis introduced by this paper states that the coefficient of variation (CV) of the annual number of foodborne illness cases decreases as result of larger number of exposures (or servings) (n). Second-order Monte-Carlo analysis and Classical statistic were used to prove the hypothesis, based on existing risk models on *L. monocytogenes* in deli meats products focused on elderly people in United States. Likewise, the hypothesis was tested on epidemiological data of annual incidence of Listeriosis in different countries (i.e. different n). Although different sources of error affected the accuracy of results, both the Monte-Carlo analysis (*in silico*) and epidemiological data (*in vivo*) demonstrated that the CV of the annual number of cases decreased as n increased as stated by the CLT. Furthermore, results from this work showed that classical statistical methods can be helpful to provide reliable risk estimations based on simple and well-established statistical principles.

Keywords: *quantitative risk analysis, predictive microbiology, Central Limit Theorem, Public Health, Monte-Carlo analysis, Food-borne diseases*

Introduction

Public Health Surveillance systems are intended to record the occurrence of diseases or intoxication as caused by pathogens and toxicants present in foods, to analyze epidemiological data and to disseminate information. The information provided by surveillance systems is crucial to design and implement strategies and/or interventions to minimize food-borne illness. However, burden-of-illness estimates often pose uncertainty due to several important limitations: test sensitivity, underreported cases, deficiencies in reporting systems, scarce human resources, etc. Despite these important sources of uncertainty, variations in the number of annual cases still remains small in comparison to the variation in pathogen doses which in most cases can span several orders of magnitude. In this work, we aimed to introduce the idea that although there may be considerable variation between individual risks, the annual variation of the total risk (number of cases) will be small as a result of the Central Limit Theorem (CLT) of probability theory. The CLT can be thought of as the cornerstone for understanding collective phenomena (Sornette *et al*. 2003). Based on the CLT and properties of variance and mean, it could be stated that the higher number of exposures (doses) to the pathogenic microorganism, the lower uncertainty on the number of attendant annual cases. From that, it could be expected that pathogenic microorganisms with high prevalence in foods such as *Listeria monocytogenes* showed a decreasing trend in the illness incidence variability as population size increases (related to number of exposures). This hypothesis was demonstrated on epidemiological data, analytical calculation and Monte-Carlo analysis.

Materials and Methods

Burden-of-illness explained by Central Limit Theorem

From a probabilistic view, the overall risk distribution can be seen as the sum of **n** individual risk distributions, being **n** the number of doses or exposures in a certain population. If **n** is

sufficiently great and none of those distributions (e.g., individual risk distributions) dominate the resultant distribution (e.g. the overall risk distribution), the CLT can be applied. The CLT states that as the number of variables increases (infinite), the sum of those variables approximates (asymptotically) to a normal distribution with parameters $\mathbf{n}^*\mu$ and $\sqrt{\mathbf{n}}^*\sigma$. The CLT applied to the sum of variables uses the properties of variance and mean to estimate μ and σ of the resultant normal distribution. It is important to note that although CLT conditions can not be reached exactly, a reasonable good approximation will be expected in a certain region around the mean whose accuracy will depend on how large the deviation from CLT is.

Based on the properties of variance and mean and CLT, it can be obtained that the coefficient of variance (CV) of the sum of variables follows the linear function:

$$\log_{10}(CV) = -0.5\log_{10}(n) + \log_{10}\left(\frac{\sigma_x}{\mu_x}\right) \quad (1)$$

Estimating Burden-of-illness by using Monte-Carlo Analysis (in silico)

Concentration at retail taken from Chen *et al*. (2003) was the initial input in the exposure assessment model previously developed by Perez-Rodriguez *et al*. (2007). After simulation, concentration at consumption in contaminated servings, i.e. doses, was obtained. The dose-response model modified by Perez-Rodriguez *et al*. (2007) was applied in the linear region ($r = 1.85\cdot10^{-14}$) and probability of getting ill (P_{ill}) was estimated by using concentration at consumption (doses) and a point-estimate value of serving size which corresponded to the mean value (64 g). To estimate the number of annual cases (i.e., overall risk), resultant distribution of individual probabilities of getting ill (individual risk) were summed **n** times by an iterative process (using Monte-Carlo analysis) being **n** the number of exposures corresponding with contaminated servings consumed by elderly population in the US which corresponded to $5.11\cdot10^7$ servings.

Estimating Burden-of-illness by applying the Central Limit Theorem

The initial input was the doses distribution taken from Monte-Carlo Analysis. Doses distribution ($\log_{10}$ cfu/serving) was described by a normal distribution with parameters σ_{logD} and μ_{logD}. The dose-response model was defined by a straight-line ($r = 1.85\ 10^{-14}$), therefore calculations of the probability of getting ill could be performed by applying the properties of variance and mean on the normal distribution of doses. If the logarithm is applied to dose-response model, then:

$$\log_{10}(P_{ill}) = \log_{10}(r) + \log_{10}(\text{Dose}) \quad (2)$$

Since variance does not change when a scalar value is summed, the distribution of $\log_{10}(P_{ill})$ denoted by $F(\text{Log}(P_{ill}))$ can be estimated according to the following expression:

$$F(\log_{10}(P_{ill})) = N(\mu_{logD} + \log_{10}(r), \sigma_{logD}) \quad (3)$$

Finally, based on CLT, the distribution of the number of cases of listeriosis can be approximated as the sum of **n** $N(\mu_D, \sigma_D)$, being **n**, the number of exposures (i.e. contaminated servings). The value for **n** was the same to that used by Monte-Carlo analysis ($\mathbf{n}=5.11\cdot10^7$):

$$F(\text{cases/year}) = N(\mathbf{n}\cdot\mu_D, \sqrt{\mathbf{n}}\cdot\sigma_D) \quad (4)$$

Epidemiological data analysis (in vivo)

Incidence data of food-borne diseases by *L. monocytogenes* in different countries around the world were collected from international and national surveillance system databases. The selection of countries was based on criteria of population size (~10^7-10^8) and reliability of the surveillance system. When possible, data were taken from the same source in order to avoid additional variation sources. Incidence data were expressed as confirmed number of cases per year. The CV was calculated for each country in the period 2002-2007 based the mean population in that same period.

Results and Discussion

Monte Carlo analysis resulted in a mean estimation of 29 annual cases of listeriosis with 95th percentile of 49 cases. On the other hand, analytical method based on the CLT obtained a lower number of cases with a mean and 95th percentile of 7 and 23 cases/year. Table 1 shows main statistics for the estimated number of listeriosis cases at a different number of exposures (**n**) for both approaches. Data revealed that both approaches converged to similar values as **n** increased. Mean values presented major similarity between both approaches at lower **n**. However, standard deviation and 95th values required a higher number of exposures (**n**) to converge. Results indicated that both approaches could be equivalent to estimate risk provided some requirements be met such as linearity in dose-response model and normality in the microbial concentration distribution as given in this example.

Table 1: Comparison between Monte-Carlo analysis and central Limit Theorem (CLT) for number of listeriosis cases at different number of contaminated servings or exposures (**n**)

n	Mean		Standard Deviation		95th		Coefficient of Variance	
	MC	CLT	MC	CLT	MC	CLT	MC	CLT
10	$4.59 \cdot 10^{-6}$	$1.47 \cdot 10^{-6}$	$2.39 \cdot 10^{-4}$	$4.13 \cdot 10^{-3}$	$1.68 \cdot 10^{-3}$	$6.80 \cdot 10^{-3}$	5.21·10	$2.80 \cdot 10^{3}$
10^2	$6.98 \cdot 10^{-5}$	$1.47 \cdot 10^{-5}$	$3.94 \cdot 10^{-3}$	$1.31 \cdot 10^{-2}$	$2.68 \cdot 10^{-4}$	$2.15 \cdot 10^{-2}$	5.65·10	$8.87 \cdot 10^{2}$
10^3	$4.72 \cdot 10^{-4}$	$1.47 \cdot 10^{-4}$	$8.71 \cdot 10^{-3}$	$4.13 \cdot 10^{-2}$	$2.23 \cdot 10^{-3}$	$6.81 \cdot 10^{-2}$	1.85·10	$2.80 \cdot 10^{2}$
10^4	$4.91 \cdot 10^{-3}$	$1.47 \cdot 10^{-3}$	$4.22 \cdot 10^{-2}$	$1.31 \cdot 10^{-1}$	$1.71 \cdot 10^{-1}$	$2.16 \cdot 10^{-1}$	8.58	8.87·10
10^5	$5.34 \cdot 10^{-2}$	$1.47 \cdot 10^{-2}$	$4.20 \cdot 10^{-1}$	$4.13 \cdot 10^{-1}$	$1.19 \cdot 10^{-1}$	$6.94 \cdot 10^{-1}$	7.86	2.80·10
10^6	$5.50 \cdot 10^{-1}$	$1.47 \cdot 10^{-1}$	1.66	1.31	1.11	2.30	3.02	8.87
10^7	5.66	1.47	8.27	4.13	9.64	8.27	1.46	2.80
*$5.11 \cdot 10^7$	2.83·10	7.53	4.22·10	9.34	4.92·10	2.29·10	1.66	1.24

*Due to computational restrictions, total number of cases for **n** = $5.11 \cdot 10^7$ were estimated by extrapolation based on the trend shown by each statistic parameter (e.g. the mean number of cases was estimated by multiplying by 5.11).

Overall, when different numbers of exposures (**n**) were studied by Monte-Carlo analysis, CV (log) reduced as **n** becomes higher (Figure 1) except for very low numbers of exposures (10-100) which did not show a clear decrease trend. Monte-Carlo analysis does not yield reliable estimations when low numbers of samples are simulated since standard deviation of the simulated distribution is not yet stabilized. Nevertheless, the discrepancies observed between CLT and Monte-Carlo analysis were progressively reduced as the number of exposures was increasing (**n** $\geq 10^6$). Convergence between both approaches is not a fact which can be observed at relatively low number of exposures because of, according to CLT, normality for sum of variables is met when **n** approximates to infinite, i.e. **n** becomes enormously high.

The regression analysis applied on Monte-Carlo analysis and Epidemiological data did not derive the exact mathematical equation given by the CLT (eq. 1). Nevertheless, regression analysis confidence intervals indicated a reasonable convergence to The Central Limit Theorem (Table 2). Probably, additional sources of uncertainty coming from the Random Number Generator seed variation and the high-dependency of Monte-Carlo analysis on the number of iterations together with the expected uncertainty derived from food-borne outbreaks reporting systems could be responsible for the lack of accuracy and precision in data.

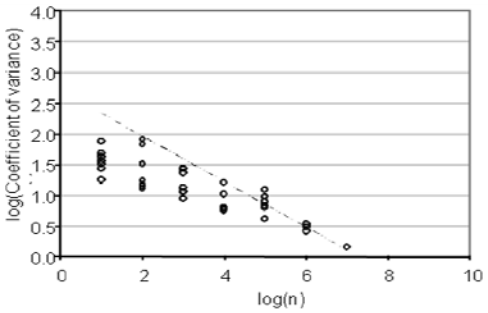

Figure 1: Representation on logarithmic scale of coefficient of variance (CV) of annual burden-of-illness obtained by Monte-Carlo analysis at different number of exposures (**n**). Solid line corresponded to the CV trend based on Central Limit Theorem (CLT) and dashed line represents the fitted eq. (1) to Monte-Carlo analysis data in range $\mathbf{n}=10^5$-10^7.

Table 2: Regression Parameters and statistics describing log-linear decrease of the coefficient of variance (eq. 1) fitted to data obtained from Monte-Carlo analysis and epidemiological data at different number of exposures (**n**).

Illness	*Data*	*m*	*Standard Error*	*p-value*	*Lower 95%*	*Upper 95%*	R^2	$R^2(m=-0.5)$
Listeriosis	Monte-Carlo	-0.37	0.06	<0.01	-0.51	-0.22	0.81	0.61
Listeriosis	Epidemiological	-0.20	0.07	0.11	-0.56	0.15	0.38	0.00

Conclusions

In many areas, the Central Limit Theorem is used as a first approach to understanding phenomena from a global perspective (e.g. economic sciences). Interpretation about reality is always complex and general rules can be helpful to extract basic and useful information. This was the main purpose in this work in which an attempt was made to study Microbiological Risk Assessment aspects from an angle of Public Health. The hypothesis suggested in this work was that "annual variation in number of food-borne illness cases is reduced as result of major exposure intensity (**n**)". Results did show a clear decreasing trend in the coefficient of variance of number of annual cases as **n** increases. Furthermore, the present study shows that classical statistical methods can be helpful to provide sound probabilistic risk estimation based on simple and well-established statistical principles.

Acknowledgements

This work was partly financed by MICINN AGL2008-03298/ALI, the Excellence Project AGR-01879 (Junta de Andalucía) and by the Research Group AGR-170 HIBRO of the "Plan Andaluz de Investigación, Desarrollo e Innovación" (PAIDI).

References

Chen Y., Ross W. H., Scott V. N. and Gombas D.E. (2003). Listeria monocytogenes: Low Levels Equal Low Risk. *Journal of Food Protection* 66, 570-577.

Pérez-Rodríguez F., van Asselt E. D., García-Gimeno R. M., Zurera G. and Zwietering M.H. (2007). Extracting additional risk managers information from a risk assessment of Listeria monocytogenes in deli meats. *Journal of Food Protection* 70, 1137–1152. 5.

Sornette, D. eds. Critical Phenomena in Natural Sciences. Chaos, Fractals, Selforganization and Disorder: Concepts and Tools. 2nd ed. New York, PA: Springer-Verlag Berlin Heidelberg; 2006.