



An algorithm to generate correlated input-parameters to be used in probabilistic sensitivity analyses

Mohamed Neine  and Desmond Curran 

GSK, Vaccines Value Evidence, Wavre, Belgium

ABSTRACT

Background: Assessment of uncertainty in cost-effectiveness analyses (CEAs) is paramount for decision-making. Probabilistic sensitivity analysis (PSA) estimates uncertainty by varying all input parameters simultaneously within predefined ranges; however, PSA often ignores correlations between parameters.

Objective: To implement an efficient algorithm that integrates parameter correlation in PSA.

Study design: An algorithm based on Cholesky decomposition was developed to generate multivariate non-normal parameter distributions for the age-dependent incidence of herpes zoster (HZ). The algorithm was implemented in an HZ CEA model and evaluated for gamma and beta distributions. The incremental cost-effectiveness ratio (ICER) and the probability of being cost-effective at a given ICER threshold were calculated for different levels of correlation. Five thousand Monte Carlo simulations were carried out.

Results: Correlation coefficients between parameters sampled from the distribution generated by the algorithm matched the desired correlations for both distribution functions. With correlations set to 0.0, 0.5, and 0.9, 90% of the simulations showed ICERs below \$25,000, \$33,000, and \$38,000 per quality-adjusted life-year (QALY), respectively, varying incidence only; and below \$38,000, \$48,000, and \$58,000 per QALY, respectively, varying most parameters.

Conclusion: Parameter correlation may impact the uncertainty of CEA results. We implemented an efficient method for generating correlated non-normal distributions for use in PSA.

ARTICLE HISTORY

Received 13 October 2020

Accepted 23 November 2020

KEYWORDS

Probabilistic sensitivity analysis; Cholesky decomposition; correlated parameter distributions; incremental cost-effectiveness analysis

Background


Cost-effectiveness analyses (CEAs) are widely used to assess the cost-effectiveness ratio of a new health technology compared to an alternative intervention [1]. Assessment of uncertainty in CEAs is paramount since it informs decision makers whether results are robust under changes in assumptions and input parameters. Uncertainties in CEAs may stem from the modeling approach (structural uncertainty) or from uncertainty in input parameters, either because of imperfect knowledge of their true value or variability due to inherent heterogeneity, for example in patient characteristics. Parameter uncertainty may virtually affect all input parameters including demographic, epidemiological, clinical, economic, and quality of life parameters. Deterministic sensitivity analysis (DSA) and probabilistic sensitivity analysis (PSA) are often used to explore the impact of input parameters on CEA outcomes [2–5].

While DSA explores to which extent CEA results change when assuming extreme values for one or

multiple parameters [6], PSA provides a global picture of uncertainty by sampling parameter values, typically using Monte Carlo (MC) simulations, from probabilistic distributions (e.g., normal, log normal, beta, gamma) [6]. In most instances, the correlation between input parameters is not incorporated in PSA [7]. Treating all input parameters as independent variables may lead to unrealistic combinations of parameter values and joint distributions that do not represent adequately the real-world situation. The importance of considering the correlation between different parameters when sampling their value for PSA has been raised in the literature [6–8]. Additionally, various guidelines on economic evaluations also recommend that the correlation among parameters should be incorporated, as it can affect both expected values and their degree of uncertainty (i.e., understating or overstating the true cost uncertainty) [6,9]. However, the techniques and algorithms for handling parameter correlation in a cost-effectiveness model, especially in the case of

CONTACT Mohamed Neine  melmoctar.neine@gmail.com  40 boulevard d'Anvers 1000 Bruxelles, Belgium

*Present address: Independent consultant

 Supplemental data for this paper can be accessed on the [here](#).

© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

multivariate non-normal distributions, are less well documented. Nevertheless, methods of generating multivariate distributions from correlated non-normal marginal distributions have been described in other fields, such as behavioral science and financial risk analysis.

In this study, we used an algorithm based on Cholesky decomposition to generate correlated multivariate non-normal distributions (e.g., beta, gamma, log-normal) for use in PSA. We evaluated the impact of correlation between input parameters on CEA results using as an example the cost-effectiveness of herpes zoster (HZ) vaccination in persons aged ≥ 50 years. Vaccination was shown to be highly cost-effective under model assumptions and PSA, but so far, the existing CEA model did not address the correlation between input parameters.

Several studies reported that HZ incidence is higher in the older age groups [10–12]. A simulation of the incidence by age group without taking into account parameter correlation may lead to inconsistent values where the incidence in older age groups may be lower than the incidence in younger age groups, thereby leading to modeling results that are inconsistent with epidemiological data.

The objectives of this study are to raise awareness of the importance of parameter correlation in CEA when evaluating uncertainty and to provide access to an algorithm that can be efficiently incorporated into existing CEA models.

Methods

Correlation of parameters

In PSA, each input-parameter is considered as a random variable, associated with a probability distribution. For each MC run, parameters are randomly sampled from their assigned distributions and the model is executed to calculate a set of outcome values (i.e., Incremental Cost-Effectiveness Ratio [ICER]). The procedure is repeated for numerous iterations, usually 1,000 or more, and the outcomes presented on a probabilistic sensitivity analysis scatter plot showing the incremental costs per quality-adjusted life-year (QALY) gained distributed across the cost-utility plane [13].

Most often, parameters are sampled independently, but this is inadequate in the presence of parameter correlation. In the case of HZ incidence, epidemiological data show that the risk of HZ increases with increasing age. The HZ CEA model used here is stratified by age, and each age group is associated with its expected HZ incidence and distribution. When sampling HZ incidence values for each age group, the correlation between sample values should be maintained to reflect the fact that HZ incidence increases with age. Therefore, these

parameters should be sampled together from a joint probability distribution of correlated parameters.

Generating a joint probability distribution from a set of uncorrelated (marginal) parameter distributions can be achieved by different approaches, such as the copula approach [8]. Here, we use a similar, well-known algorithm, i.e., the Cholesky decomposition matrix, which allows introducing pre-specified correlation among a set of uncorrelated variables.

Let us assume that the variables $X_1, X_2, \text{ and } X_3$ are correlated using the correlation matrix:

$$C = \begin{bmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{bmatrix}$$

with an associated Cholesky decomposition matrix L , such that $C = L^T L$ [14,15].

$$L = \begin{bmatrix} 1 & 0.5 & 0.5 \\ 0 & 0.866 & 0.2887 \\ 0 & 0 & 0.8165 \end{bmatrix}$$

If we assume that vector Y represents the three independent variables $Y_1, Y_2, \text{ and } Y_3$, then the vector $X = YL$ represents three correlated variables $X_1, X_2, \text{ and } X_3$ with correlation matrix C [14,15].

Sampling algorithm

The matrix L is known as the Cholesky decomposition matrix of the correlation matrix C . To simplify, we assumed that the correlation values between N parameters are equal ($=\rho$). We then simulated the correlated multivariate non-normal distribution following these steps:

- (1) We created a Cholesky decomposition matrix using the correlation value ρ and the number of parameters correlated (N);
- (2) We simulated N uncorrelated normal random variables with a *mean* $\mu = 0$ and a *standard deviation* $\sigma = 1$;
- (3) We multiplied the uncorrelated normal variables by the Cholesky matrix to generate correlated normal random variables and calculated the probability associated with each value from the cumulative distribution function from a normal distribution;
- (4) We calculated the inverse probability of these values assuming a non-normal distribution (e.g., beta, gamma, log-normal).

The algorithm was implemented using visual basic for applications (VBA) in Microsoft Excel (2007) and further explanations are provided in the user guide (Additional file 1).

Implementation

The algorithm was implemented in a static multi-cohort Markov model developed to assess the cost-effectiveness of an adjuvanted recombinant zoster vaccine (RZV) vs no vaccination in adults aged ≥ 50 years [16]. Input data used in the model were retrieved from the RZV cost-effectiveness study recently published [17]. Cohorts were split into five age groups (i.e., 50–59, 60–64, 65–69, 70–79, and ≥ 80). The CEA was performed from a US societal perspective over a lifetime horizon. The primary outcome in the model was the ICER. The model results and ICERs including uncertainty were estimated by carrying out two PSAs consisting of 5,000 simulations each. In the first PSA, only HZ incidence was varied using multivariate distributions of correlated parameters generated by Cholesky decomposition. In the second PSA, besides HZ incidence most other input parameters were varied, except for all-cause mortality, vaccine price per dose, administration cost per dose, number of doses, the time between first and second dose, coverage of first dose vaccination, number of doses and discounting.

To assess the performance of the Cholesky decomposition algorithm in generating joint distributions of variables with a pre-specified correlation, we first simulated sets of parameters where the correlation was set to the desired values of 0.25, 0.5, 0.75, and 0.9. Next, we evaluated the impact of parameter correlation on PSA results. Two PSAs were carried out as described above: in the first PSA only correlation between HZ incidence parameters was considered, while in the second PSA, most model parameters were treated as correlated parameters. Furthermore, three different scenarios were evaluated, depending on the desired correlation values:

Zero correlation: no correlation, parameters are assumed to be independent ($\rho = 0.0$);

Moderate correlation ($\rho = 0.5$);

Strong correlation ($\rho = 0.9$).

Key parameters in the model and the distribution associated with each parameter in the PSA are presented in Additional file 2. Gamma distributions were assumed for all cost variables while beta distributions were applied for all other inputs [4,5]. The results of the PSA were presented separately on a cost-effectiveness plane and a cost-effectiveness acceptability curve.

Results

The Cholesky decomposition matrix has been embedded in a Microsoft Excel workbook via a VBA macro and is provided in Additional file1. We used a 5×5 matrix to match the number of age groups

(i.e., 50–59, 60–64, 65–69, 70–79, and ≥ 80) used in our CEA example of HZ vaccination. The user can integrate this tool in any Excel-based PSA project to take into account the correlation between a set of correlated parameters. The correlation coefficient and type of distribution can be adapted by the user.

Adding the correlation between selected parameters using Cholesky decomposition added very little computational overhead. The actual runtime will depend on the CEA model and the computational platform.

Simulation results

Cholesky decomposition generated multivariate parameter distributions with correlations between age-dependent HZ incidence values close to the desired correlations, regardless of the underlying marginal distribution (beta or gamma distribution). The median correlation estimated through the simulation assuming a gamma distribution was 0.247, 0.496, 0.748, and 0.899 for the desired correlations of 0.25, 0.5, 0.75, and 0.9, respectively. Similarly, the median correlation estimated assuming a beta distribution was 0.248, 0.494, 0.746, and 0.895, respectively (Additional files 3 and 4).

Cost-effectiveness results

The base case results (e.g., the results of the CEA model) showed that RZV was associated with an ICER estimated at 11,863 USD per quality-adjusted life-year (QALY) gained compared to no vaccination.

For the PSAs varying incidence only, 90% of the simulations resulted in an ICER below 25,000 USD, 33,000 USD, and 38,000 USD using a correlation of 0.0, 0.5, and 0.9, respectively (Figure 1(a)). For the PSAs varying most parameters in the model, 90% of the simulations resulted in an ICER below 38,000 USD, 48,000 USD, and 58,000 USD using a correlation of 0.0, 0.5, and 0.9, respectively (Figure 1(b)).

Uncertainty around the incremental QALYs and incremental costs for the two scenarios is presented by scatterplots (Figure 2(a,b)). The QALYs gained ranged from approximately 1200 to 3600 while the costs varied from approximately 100 million in expenditure to savings of 145 million per 1 million individuals vaccinated. As expected, uncertainty was larger in scenario 2 in which most parameters were varied compared with scenario 1, in which only a subset of parameters related to HZ incidence was varied. Including correlation between the age-dependent HZ incidence also tended to increase uncertainty as illustrated by dots lying on the outside of the scatterplot cloud.

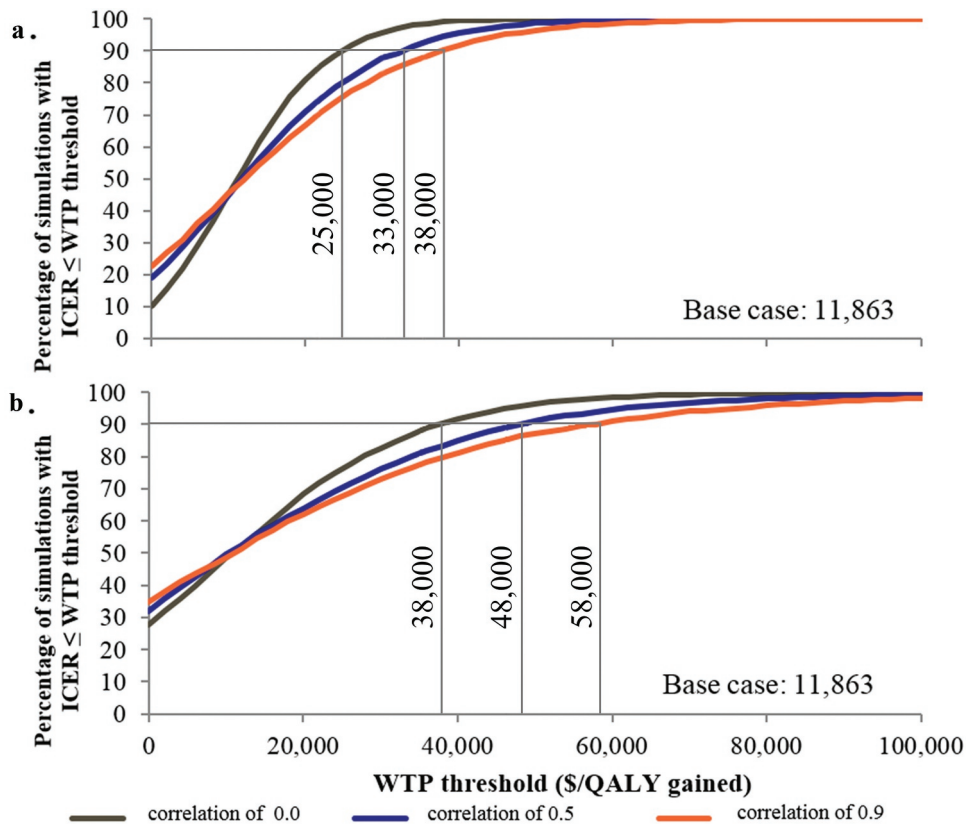


Figure 1. Cost-effectiveness acceptability curves varying (a) incidence parameters only; or (b) most parameters.

Acceptability curves with a threshold of 90% of the simulations being cost-effective for PSA. ICER, incremental cost-effectiveness ratio; PSA, probabilistic sensitivity analysis; QALY, quality-adjusted life-year; WTP, willingness-to-pay.

Discussion

PSAs provide a synthesis of parameter uncertainty and its impact on the cost-effectiveness of a given intervention; guidelines recommend the use of PSA to assess uncertainty in CEA and PSAs have been widely used in health economics evaluations [4]. However, a review of CEAs conducted for the National Institute for Health and Clinical Excellence (NICE) revealed that the majority of models ignore the potential correlation between input parameters [7]. The apparent lack of considering parameter correlation in CEAs might be due to the unfamiliarity of the analyst with the subject, programming issues, and the absence of explicit requirements and recommendations in applicable guidelines. The goal of this study was to overcome these hurdles by describing a known algorithm for handling parameter correlation and providing access to its implementation in Microsoft Excel, a program widely used for CEA modeling. Similar efforts have been undertaken by several other groups, using different methods to introduce correlation in PSA, with accompanying code provided for the statistical package R [8,18].

Our algorithm used a Cholesky decomposition to transform independently correlated normal random variables into correlated normal random variables. These variables were then transformed into non-normal variables resulting in correlated non-normal random variables. We simulated 5,000 values for each parameter included in the PSA; after 5,000 simulations, the results appeared to be stable and no additional simulations were needed [19]. The results showed that the desired correlations were met by the algorithm, as the median values were close to the desired correlations for both distributions. The scatter plot clouds produced by varying only the incidence parameters are more densely distributed than when most parameters are varied, which reflects the impact of the different parameters on the PSA analysis.

In our CEA model of HZ vaccination, including the (positive) correlation between age-dependent HZ incidence lead to higher uncertainty in CEA outcomes, i.e., incremental costs and incremental QALYs. As expected, the credibility interval changed dramatically when more parameters were introduced in the PSA. Other situations may occur where consideration of the correlation between input parameters may decrease the uncertainty

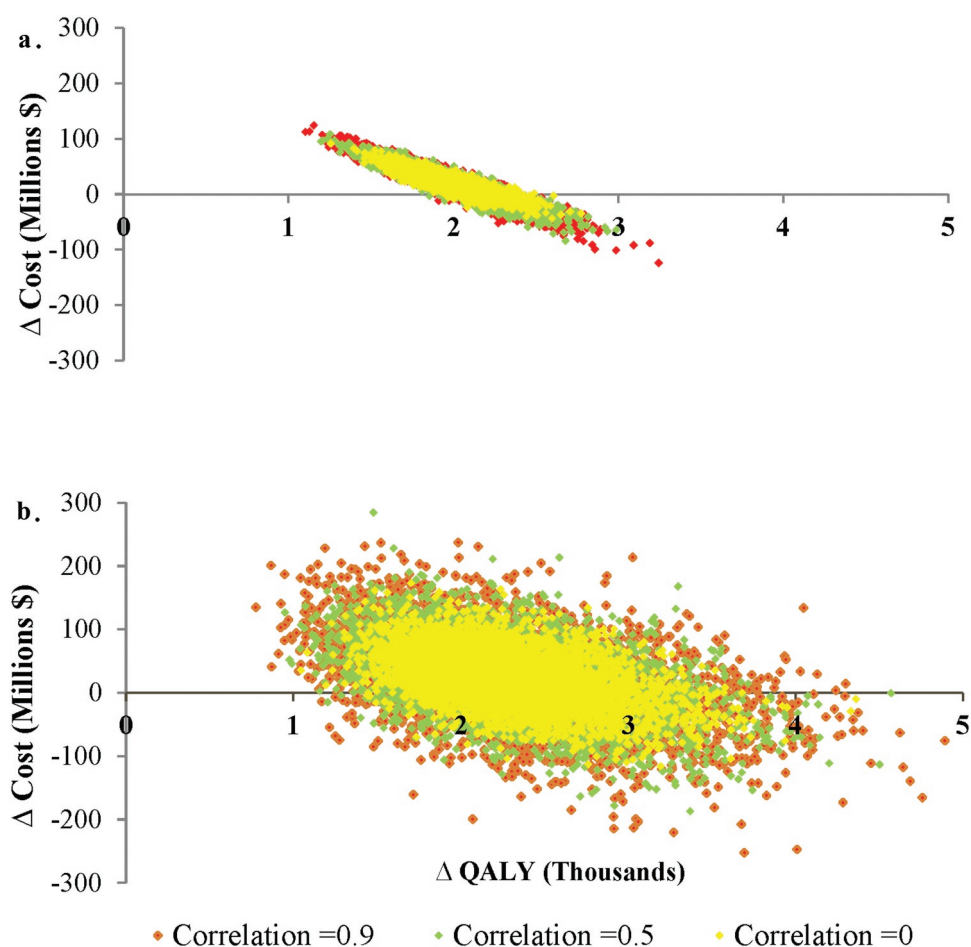


Figure 2. ICER scatterplot of scenarios for PSA varying (a) incidence parameters only; or (b) most parameters. PSA, probabilistic sensitivity analysis; QALY, quality-adjusted life-year; ICER, incremental cost-effectiveness ratio.

in cost-effectiveness results [8]. In either case, the impact of parameter correlation provides a more realistic estimate of the uncertainty in CEA results. PSA may also highlight areas where additional information might be valuable to reduce uncertainty and thereby lending more credibility to model outcomes.

There are several limitations to the presented method. The results provided by this algorithm are imperfect in terms of the desired correlations and the correlations obtained based on simulations, e.g., for the desired correlation of 0.25 and assuming a beta distribution, the median was 0.247 with a range of 0.184 to 0.277 (Additional file 3). The skewed nature and ranges observed are similar to those presented by Iman and Conover 1982 [14].

Frequently, the true value of a correlation is unknown. Ideally, all available evidence, including expert opinion, should be used to estimate uncertainty both around input parameter values and correlation. While the presented method allows for propagating uncertainty in input parameters to assess uncertainty in CEA output variables, the correlation between input variables is considered as a fixed variable in the PSA. Other methods of simulating

multivariate non-normal variables have been proposed which are more robust and allow for considering uncertainty in correlation itself, but these methods require iterative processes and are consequently less efficient from a computational perspective [6,18]. The described method is intended for situations where the exact correlation remains unknown, but where several plausible assumptions can be explored. Another consideration is the potential non-linearity of relationships, which may influence model calculations [8].

In conclusion, considering correlations in PSA may impact conclusions reached based on CEA results. Analysis of decisions taken by NICE on a series of CEA models suggests that the outcome depended on uncertainty and that high uncertainty was more likely leading to a negative decision [7]. Although the precise value of parameter correlation is often unknown, allowing for parameter correlation will help decision makers understand the potential impact on results and overall model uncertainty. Here, we developed an algorithm based on Cholesky decomposition, which simulates a correlated non-normal distribution taking any desired correlation

value. The VBA code can be freely downloaded from the link attached to this paper.

Acknowledgments

The authors would like to thank Stephane Lorenc (GSK) and Desiree Van Oorschot (GSK) for their contribution to the study. The authors also thank the Business & Decision Life Sciences platform for editorial assistance, manuscript coordination and writing support, on behalf of GSK. Gregory Collet and Bruno Dumont coordinated manuscript development and editorial support.

Disclosure statement

DC is an employee of the GSK group of companies and holds shares in the GSK group of companies. MN was an employee of the GSK group of companies at the time of the study. DC and MN declare no other financial and non-financial relationships and activities.

Funding

GlaxoSmithKline Biologicals SA funded this study (GSK study identifier: HO-13-13333) and was involved in all stages of study conduct, including analysis of the data. GlaxoSmithKline Biologicals SA also took charge of all costs associated with the development and publication of this manuscript.

Availability of data and material

GSK makes available anonymized individual participant data and associated documents from interventional clinical studies that evaluate medicines, upon approval of proposals submitted to www.clinicalstudydatarequest.com. To access data for other types of GSK sponsored research, for study documents without patient-level data, and for clinical studies not listed, please submit an inquiry via the website.

Authors' contributions

All authors participated in the conception and design of the study, participated in the collection and generation of the study data, performed the analysis, and were involved in the interpretation of the data. All authors had full access to the data and the corresponding author was responsible for submission of the publication.

ORCID

Mohamed Neine  <http://orcid.org/0000-0003-2980-4232>
Desmond Curran  <http://orcid.org/0000-0002-7423-0111>

References

- [1] Hjelmgren J, Berggren F, Andersson F. Health economic guidelines-similarities, differences and some implications. *Value Health*. 2001;4:225–250.
- [2] Adalsteinnsson E, Toumi M. Benefits of probabilistic sensitivity analysis - a review of NICE decisions. *J Mark Access Health Policy*. 2013;1.
- [3] Claxton K, Martin S, Soares M, et al. Methods for the estimation of the national institute for health and care excellence cost-effectiveness threshold. *Health Technol Assess*. 2015;19(1–503):v–vi. .
- [4] Briggs AH, Weinstein MC, Fenwick EA, et al. Model parameter estimation and uncertainty: a report of the ISPOR-SMDM modeling good research practices task force–6. *Value Health*. 2012;15:835–842.
- [5] Briggs AH, Goeree R, Blackhouse G, et al. Probabilistic analysis of cost-effectiveness models: choosing between treatment strategies for gastroesophageal reflux disease. *Med Decis Making*. 2002;22:290–308.
- [6] Ades AE, Claxton K, Sculpher M. Evidence synthesis, parameter correlation and probabilistic sensitivity analysis. *Health Econ*. 2006;15:373–381.
- [7] Andronis L, Barton P, Bryan S. Sensitivity analysis in economic evaluation: an audit of NICE current practice and a review of its use and value in decision-making. *Health Technol Assess*. 2009;13:iii, ix–xi, 1–61.
- [8] Naversnik K, Rojnik K. Handling input correlations in pharmacoeconomic models. *Value Health*. 2012;15:540–549.
- [9] CADTH methods and guidelines. Guidelines for the economic evaluation of health technologies: Canada. 4th ed. Ottawa, Canada; 2017. <https://www.cadth.ca/about-cadth/how-we-do-it/methods-and-guidelines/guidelines-for-the-economic-evaluation-of-health-technologies-canada>.
- [10] Johnson BH, Palmer L, Gatwood J, et al. Annual incidence rates of herpes zoster among an immunocompetent population in the USA. *BMC Infect Dis*. 2015;15:502.
- [11] Scott FT, Johnson RW, Leedham-Green M, et al. The burden of herpes zoster: a prospective population based study. *Vaccine*. 2006;24:1308–1314.
- [12] Yawn BP, Saddier P, Wollan PC, et al. A population-based study of the incidence and complication rates of herpes zoster before zoster vaccine introduction. *Mayo Clin Proc*. 2007;82:1341–1349.
- [13] O'Hagan A, Stevenson M, Madan J. Monte Carlo probabilistic sensitivity analysis for patient level simulation models: efficient estimation of mean and variance using ANOVA. *Health Econ*. 2007;16:1009–1023.
- [14] Iman RL, Conover WJ. A distribution-free approach to inducing rank correlation among input variables. *Commun Stat - Simul Comput*. 1982;11:311–334.
- [15] Scheuer EM, Stoller DS. On the generation of normal random vectors. *Technometrics*. 1962;4:278–281.
- [16] Curran D, Van Oorschot D, Varghese L, et al. Assessment of the potential public health impact of Herpes Zoster vaccination in Germany. *Hum Vaccin Immunother*. 2017;13:2213–2221.
- [17] Curran D, Patterson B, Varghese L, et al. Cost-effectiveness of an adjuvanted recombinant zoster vaccine in older adults in the USA. *Vaccine*. 2018;36:5037–5045.
- [18] Ruscio J, Kaczetow W. Simulating multivariate nonnormal data using an iterative algorithm. *Multivariate Behav Res*. 2008;43:355–381.
- [19] Hattswell AJ, Bullement A, Briggs A, et al. Probabilistic sensitivity analysis in cost-effectiveness models: determining model convergence in cohort models. *Pharmacoeconomics*. 2018;36(12):1421–1426. .