VALUE IN HEALTH ▮ (2017) ▮▮▮–▮▮▮

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jval

A Practical ANOVA Approach for Uncertainty Analysis in Population-Based Disease Microsimulation Models

Behnam Sharif, PhD[1,], Hubert Wong, PhD[2], Aslam H. Anis, PhD[2], Jacek A. Kopec, PhD[2]*

[1]*Faculty of Medicine, Department of Community Health Sciences, University of Calgary, Calgary, Alberta, Canada;* [2]*School of Population and Public Health, University of British Columbia, Vancouver, British Columbia, Canada*

ABSTRACT

Objectives: To provide a practical approach for calculating uncertainty intervals and variance components associated with initial-condition and dynamic-equation parameters in computationally expensive population-based disease microsimulation models. **Methods:** In the proposed uncertainty analysis approach, we calculated the required computational time and the number of runs given a user-defined error bound on the variance of the grand mean. The equations for optimal sample sizes were derived by minimizing the variance of the grand mean using initial estimates for variance components. Finally, analysis of variance estimators were used to calculate unbiased variance estimates. **Results:** To illustrate the proposed approach, we performed uncertainty analysis to estimate the uncertainty associated with total direct cost of osteoarthritis in Canada from 2010 to 2031 according to a previously published population health microsimulation model of osteoarthritis. We first calculated crude estimates for initial-population sampling and dynamic-equation parameters uncertainty by performing a small number of runs. We then calculated the optimal sample sizes and finally derived 95% uncertainty intervals of the total cost and unbiased estimates for variance components. According to our results, the contribution of dynamic-equation parameter uncertainty to the overall variance was higher than that of initial parameter sampling uncertainty throughout the study period. **Conclusions:** The proposed analysis of variance approach provides the uncertainty intervals for the mean outcome in addition to unbiased estimates for each source of uncertainty. The contributions of each source of uncertainty can then be compared with each other for validation purposes so as to improve the model accuracy.

Keywords: methodology, microsimulation, probabilistic sensitivity analysis, uncertainty analysis.

Introduction

Population-based disease microsimulation (PDMS) models are individual-level disease simulation models integrated with validated demographic estimates [1]. These models incorporate statistical models to reflect past and future birth and immigration and mortality rates in addition to modeling health determinants associated with a disease and its burden across a population [1–3]. Over the past decade, there has been a growing interest in the development of chronic disease PDMS models [2] capable of projecting the distributions of health determinants and diseases in the population and predicting the effects of disease prevention and treatment strategies on health outcomes [1–3]. Examples include PDMS models of breast cancer [4], stroke [5], pulmonary disease [6], colon cancer [7], diabetes [8], and osteoarthritis [2]. Nevertheless, to provide inference on population health measures, PDMS models need to be equipped with tools to estimate the uncertainty associated with the outcome [9,10].

Uncertainty analysis (UA) in simulation models is the study of variability in the outcome that results from the propagation of uncertainties associated with model inputs [9]. The sources of uncertainty in PDMS include first-order (stochastic or aleatory), second-order (epistemic or parameter), and third-order (structural) uncertainty [9,10]. Although the parameter uncertainty reflects the lack of complete knowledge about the parameters, the first-order uncertainty reveals the nature of uncertain events inside the simulation model [9,11,12]. The third-order uncertainty is the structural uncertainty or the model uncertainty that is associated with alternative data sets and statistical models used in the model [13,14].

We evaluated uncertainty associated with two groups of parameters, initial-condition and dynamic-equation parameters, in addition to the first-order uncertainty; we did not include the structural uncertainty in this study. Similar to system dynamics [15], a major type of parameter uncertainty that pertains to microsimulation models is uncertainty in the initial-condition parameters [14]. In PDMS, this type of uncertainty reflects the variability of the outcome because of the uncertainty about the parameters used to estimate the characteristics of the individuals at baseline [3]. The remaining parameters in PDMS represent

* *Address correspondence to:* Behnam Sharif, Faculty of Medicine, Department of Community Health Sciences, University of Calgary, HRIC 3C64, 3280 Hospital Drive NW, Calgary, Alberta, Canada T2N 4Z6.

E-mail: behnam.sharif@ucalgary.ca.

1098-3015$36.00 – see front matter

http://dx.doi.org/10.1016/j.jval.2017.01.002

dynamic equations that govern the internal relations between model variables and outcomes, such as health determinants (risk factors), disease occurrence, quality of life, and cost [16]. Separating initial-condition uncertainty from dynamic-equation uncertainty would aid in providing evidence for the validation of models used to estimate each set of parameters [1].

Given the computationally expensive nature of PDMS models [11,17], performing UA with the standard Monte-Carlo (MC) method is often unmanageable. In models with two levels of uncertainty, that is, parameter and first-order uncertainty, the standard practice of probabilistic sensitivity analysis (PSA) is to apply the MC method by probabilistically sampling from the distribution of parameters in addition to using large numbers of simulated individuals. With the aim of isolating the uncertainty associated with two sets of parameters, an extra looping within the MC method is needed, which significantly increases the number of simulation runs required. Recent studies within the health economics literature have addressed the issue of reducing the computational time for UA in individual-level simulation models using Gaussian processes [13,18] and approaches based on analysis of variance (ANOVA) [17,18].

In this study, we describe ANOVA-based estimators for the variance components in the simulation outcome associated with three sources of uncertainty (initial-condition parameters, dynamic-equation parameters, and first order) and derive formulae for the numbers of simulation replicates required for each source of uncertainty to attain the minimum error given a fixed computational time. We then apply the proposed approach to estimate the uncertainty intervals associated with the total direct cost of osteoarthritis (OA) in Canada from 2010 to 2031 as projected by a previously published population health model of osteoarthritis (POHEM-OA), an individual-level continuous-time microsimulation model [2,19].

Methods

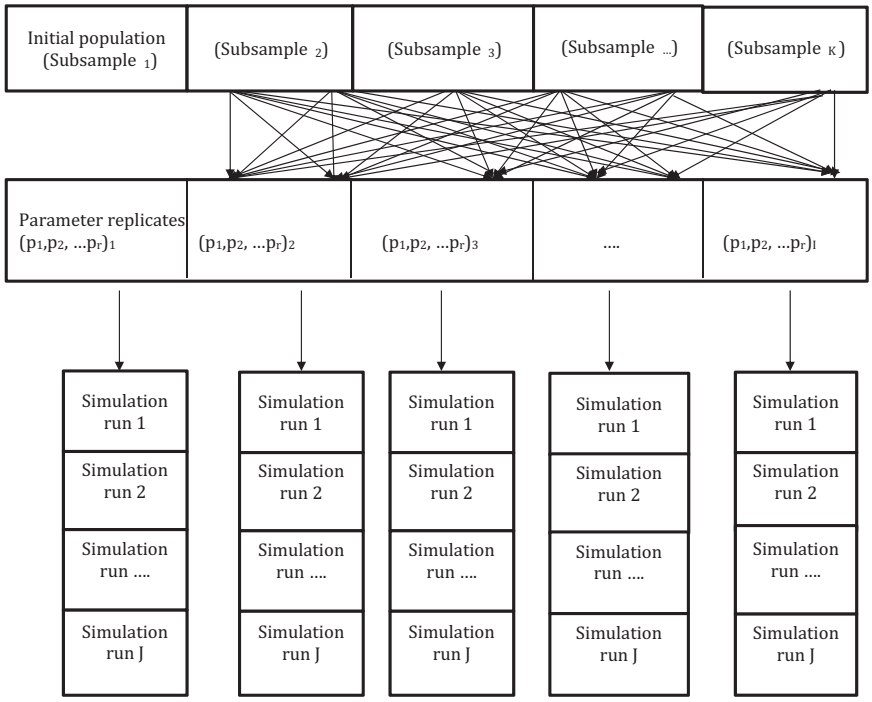
The UA procedure will be described in two stages: 1) the sampling stage, in which samples are drawn from the probability distributions of the uncertainty sources and the simulation outcome is calculated for each sample, and 2) the analysis stage, in which variance of the mean outcome and variance components associated with each source of uncertainty are estimated.

Sampling Stage

Let a_k ($k = 1, \ldots, K$) denote a random sample of K replicates from the parameters associated with initial conditions and b_j ($j = 1, \ldots, J$) denote a random sample of J replicates from the parameters associated with dynamic equations. For each of the KJ combinations of completely crossed levels of these two factors, we perform I simulation runs (replicates) reflecting first-order uncertainty and obtain the simulation outcome z_{kji} (Fig. 1). In population-based microsimulation models, initial conditions refer to initial values for characteristics of individuals such as age, sex, health determinants, and disease risk factors. In PDMS models, the initial values can be estimated using health data sources or they can be imputed using population-level health surveys. In our setting, sample replicates to evaluate uncertainty associated with initial conditions are referred to as "initial-population subsamples" and replicates associated with dynamic equations are referred to as "parameter replicates."

Analysis Stage

After the sampling step is performed, the simulated data are analyzed to estimate variance components associated with each



**Sampling procedure for two-level ANOVA model with interaction. As shown above, every initial population subsample can be crossed with every parameter replicates drawn from a multivariate distribution of parameter space.

Fig. 1 – Sampling procedure for a two-way crossed design.