

Research Article

Uncertainty Analysis in Population-Based Disease Microsimulation Models

Behnam Sharif,^{1,2} Jacek A. Kopec,^{1,2} Hubert Wong,^{1,3} Philippe Finès,⁴ Eric C. Sayre,² Ran R. Liu,² and Michael C. Wolfson^{4,5}

¹ School of Population and Public Health, University of British Columbia, 2329 West Mall, Vancouver, BC, Canada V6T 1Z4

² Arthritis Research Centre of Canada, 895 10th Avenue West, Vancouver, BC, Canada V5Z 1L7

³ Methodology & Statistics, CIHR Canadian HIV Trials Network, St. Paul's Hospital, No. 620, 1081 Burrard Street, Vancouver, BC, Canada V6Z 1Y6

⁴ Health Analysis Division, Statistics Canada, 150 Tunney's Pasture Driveway, Ottawa, ON, Canada K1A 0T6

⁵ Faculty of Medicine, University of Ottawa, 550 Cumberland Street Ottawa, ON, Canada K1N 6N5

Correspondence should be addressed to Behnam Sharif, behnam.sharif@gmail.com

Received 27 February 2012; Revised 9 June 2012; Accepted 12 June 2012

Academic Editor: Carolyn Rutter

Copyright © 2012 Behnam Sharif et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Objective. Uncertainty analysis (UA) is an important part of simulation model validation. However, literature is imprecise as to how UA should be performed in the context of population-based microsimulation (PMS) models. In this expository paper, we discuss a practical approach to UA for such models. *Methods.* By adapting common concepts from published UA guidelines, we developed a comprehensive, step-by-step approach to UA in PMS models, including sample size calculation to reduce the computational time. As an illustration, we performed UA for POHEM-OA, a microsimulation model of osteoarthritis (OA) in Canada. *Results.* The resulting sample size of the simulated population was 500,000 and the number of Monte Carlo (MC) runs was 785 for 12-hour computational time. The estimated 95% uncertainty intervals for the prevalence of OA in Canada in 2021 were 0.09 to 0.18 for men and 0.15 to 0.23 for women. The uncertainty surrounding the sex-specific prevalence of OA increased over time. *Conclusion.* The proposed approach to UA considers the challenges specific to PMS models, such as selection of parameters and calculation of MC runs and population size to reduce computational burden. Our example of UA shows that the proposed approach is feasible. Estimation of uncertainty intervals should become a standard practice in the reporting of results from PMS models.

1. Introduction

Computer simulation models are widely used in public health research [1, 2]. Population-based microsimulation (PMS) models are increasingly used to model possible effects of public health interventions at the population level [3–5]. Such models usually represent the population of a country: incorporate multiple cohorts, and model births, deaths, and migration [6–8]. Population-based models differ from models commonly used in cohort-based cost-effectiveness studies that model a single cohort of patients [9]. Unlike macrolevel simulation models (e.g., cell-based [10] or compartmental models [11]), microsimulation (MS) models generate a life history for every individual in a population [12, 13] and provide population-level outcomes by aggregating the individuals' event histories [14, 15].

In PMS models of chronic, noncommunicable diseases, individuals can be treated as independent units (no inter-individual interactions). Examples include models of breast cancer [7, 16], stroke [6, 17], pulmonary disease [18], colon cancer [19], diabetes [20, 21], and other chronic conditions [5, 15]. MS models that incorporate interactions between individuals, often referred to as agent-based models, have been used to model infectious diseases, such as influenza [22] and HIV [23]. The focus of the current paper is on non-interactive PMS models of chronic diseases.

Uncertainty analysis (UA) plays an important role in the validation of simulation models and interpretation of their results [13, 24, 25]. The purpose of UA is to provide uncertainty intervals around the mean estimate of one or more outcomes [24, 26]. In UA, the model analyst attempts to quantify the uncertainty around the outcomes that is

propagated through the model from different sources of uncertainty [24–30]. These sources include Monte Carlo (MC) error (i.e., random error), parameter uncertainty (i.e., uncertainty resulting from the fact that parameters used in simulation models are estimated and not truly known [28, 29], structural uncertainty, that is, the uncertainty associated with the choice of the statistical models [31], and possibly other sources, such as the choice of the starting population and sources of data that inform the model [25, 27].

In this paper, the term “parameter” refers to all possible model inputs. Examples of parameters in PMS models are those used to define the probability of events, such as disease onset, death, or birth, or to describe the probability of changes in the risk factors at the individual level, such as changes in body mass index (BMI) or blood pressure. Parameters can be classified into those derived directly from data and those obtained from experts. Data-based parameters can be further categorized into those estimated from a sample and, therefore, subject to sampling uncertainty, and those obtained from a population census (no sampling uncertainty). Expert-derived parameters are characterized by subjective uncertainty [13]. Further in the text, we will discuss how to deal with each type of parameter in UA.

Several methods have been developed for capturing uncertainty in the analysis of environmental risk assessment models [32], cost-effectiveness models in health economics [33], and technology assessment models [34]. In environmental risk assessment models [32], prediction of quantities such as risks, exposures, and their effect on human health is critical for public policy decision making. Prediction is also a major objective of PMS disease models [1, 8]. Providing the distribution for the outcomes or uncertainty intervals around the outcome estimates is an important step in ascertaining predictive validity of surveillance and risk assessment models [30]. In cost-effectiveness decision analytic models [33], as opposed to predictive risk assessment models, policy decisions are made through comparing cost and effectiveness of multiple scenarios by predefined objective functions, such as incremental cost-effectiveness ratio (ICER) or incremental net benefit [33, 34]. An important difference between these two types of models with regards to UA is that in predictive models the goal is to estimate the uncertainty around a projected health outcome under a given scenario, while in decision analytic models one is usually interested in the uncertainty around a relative measure comparing different scenarios [35, 36].

Probabilistic sensitivity analysis (PSA) is the term used for UA in decision analytic models [34], while quantitative UA is the term used in risk assessment and environmental modeling [32]. UA is considered a necessary component of a model-based analysis in almost all risk assessment modeling guidelines [32] and recent decision analytic modeling guidelines [34]. For instance, sensitivity analysis (deterministic or probabilistic) was cited as a requirement in the principles of good practice for decision analytical modeling in healthcare evaluations, developed by a task force from the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) and published in 2003 [37]. UA is also required for health technology assessment modeling according to the

guidelines provided by the National Institute for Health and Clinical Excellence (NICE) in the UK [34]. Performing UA is important because in certain situations, it may change the decision regarding the optimal scenario [35]. For example, Griffin et al. [29] showed that in complex MS models with nonlinear relations between the variables, interpretation of the outcomes such as ICER could be biased if uncertainty around the estimates is not considered.

UA is most informative when applied to complex models that combine data from different sources [4–7]. In addition to providing uncertainty intervals around the outcomes, UA is also used as a prerequisite for value of information analysis in decision analytic models [34] or stochastic sensitivity analysis (SA) in environmental models [37]. In both of these approaches, the goal is to ascertain the contribution of different sources of uncertainty (e.g., different parameters) to the total uncertainty of the model [22, 33, 36, 37]. As a result, critical regions of the parameter space can be located and future research can be prioritized to obtain better estimates of the key parameters and reduce the uncertainty [27].

Differences in the structure and intended applications between PMS models and other types of simulation models may affect the methodology of UA. As the sources of uncertainty in individual-level models differ from those in macrolevel models, the UA methods developed for macrolevel models [38] should be modified for use in PMS models. Studies by O’Hagan et al. [39] discuss UA for individual-level cost-effectiveness models. However, their approaches are not suited for models used for predictive purposes, such as PMS models of chronic diseases used in health impact assessment studies [1, 6, 7], or those used for population projections [26]. A guideline for conducting UA in the context of predictive models for risk assessment studies has been published by the Environmental Protection Agency (EPA) [40], but this guideline does not address the specific problems that may arise while performing UA for PMS models. In sum, no standard approach for UA in the context of PMS models has been developed [13, 25].

The uncertainty associated with any given outcome in a disease simulation model is typically a reflection of the complex interplay among the model variables [12]. Due to this complexity, estimating uncertainty with analytical approaches is often impossible, and numerical methods, such as the Monte Carlo (MC) method, are almost always required. The MC method involves running the model many times using randomly selected samples from the input parameter space (i.e., parameter joint distribution) [32]. While the MC method is the most common numeric approach to estimating parameter uncertainty [30, 32] in simulation models, in its simplest (standard) form it is not suitable for MS models due to the resulting computational burden [35].

Various techniques have been proposed to reduce the computational time of UA in MS models. Cronin et al. [31] used a response surface approach, which approximates the simulation outcome as a linear function of the input parameters. Others employed more complex approximate functions, such as Gaussian and radial basis functions [41, 42]. The major limitation of these techniques is that the assumptions of a specific response surface function may not

hold in all types of MS models or even for different outcomes within one model [42]. Another approach is to perform “guided” sampling, for example Latin hypercube sampling (LHS) [31, 43] or orthogonal sampling [44]. However, these sampling approaches do not reduce the model execution time, which is the “bottleneck” of the total computational time in UA [45]. O’Hagan et al. presented a method based on the analysis of variance and discussed sample size calculations for UA in MS models [39].

In this paper we provide an overview of the main issues and techniques in UA that are applicable to MS models. We illustrate these techniques with an example in which UA is performed for POHEM-OA [46], a PMS model of osteoarthritis in Canada.

2. Overview of Uncertainty Analysis

2.1. Sources of Uncertainty. The sources of uncertainty vary with the type of simulation model and the objectives and design of a given study. Several sources have been discussed in the literature [30, 44], with the emphasis on (i) MC error, or first-order uncertainty [2, 29, 36], and (ii) parameter uncertainty, or second-order uncertainty [38, 39, 42]. MC error is introduced when the starting values of the variables in the model and the simulated events are assigned to each individual using a stream of random numbers [9]. As such, MC error reflects “unexplained” variation between the units of a simulation model [32]. Parameter uncertainty arises from random errors in the estimation of the input parameters [8, 30]. Parameter uncertainty exists in all manners of simulation models. MC error, on the other hand, is not a concern in macrolevel models, which do not attempt to reflect stochastic variability between individuals [47].

Other types of uncertainty, including uncertainty in the starting population [12] and structural model uncertainty [16], have been studied. Structural uncertainty is associated with the choice of model structure, including the statistical models used, variables in the model, and the specified relationships between the variables [31]. Examples of structural uncertainty commonly discussed in the literature are uncertainty arising from alternative model types, such as linear versus nonlinear dose response in environmental risk assessment models [30] or the use of discrete event versus state-based Markov models [16]. Structural uncertainty is not considered further in the UA approach discussed in this paper.

2.2. Uncertainty Analysis of Microsimulation Models. UA in individual-level simulations needs to account for MC error in addition to parameter uncertainty. This can be accomplished by performing a nested analysis that iteratively samples from the parameter space, calls the base model (which itself is subject to MC error as it samples individuals) with a new set of input parameters, and records the outcomes [37]. Throughout this paper, this approach is referred to as “the MC method” in the context of PMS models. It should be noted that MC error is reducible by increasing the number of simulated units or the number of simulation runs with

different random seeds. However, parameter uncertainty is fixed and cannot be reduced [39].

There are two major issues when applying the MC method to PMS models: (i) integrating MC and calibration algorithm(s) inside the simulation model and (ii) reducing the computational burden associated with the MC method. Calibration algorithms are used extensively in PMS models [13]. Ideally, the calibration algorithm should be integrated into the UA, as discussed in Section 3.4. A Bayesian approach that integrates calibration algorithms within the MC method has been recently applied in MS disease models for estimating unknown parameters [48].

2.3. Computational Time in Uncertainty Analysis. In MS models, a major burden in applying the MC method for UA is the computational time. UA should be performed using an outer loop of (n) model runs for the parameter uncertainty and an inner loop for a population of (m) units for the MC error [36]. To obtain an accurate estimate of parameter uncertainty in macrolevel models, it has been suggested that the number of model runs (n) (i.e., MC runs) should be large enough so that the uncertainty estimates converge [29]. However, applying this recommendation to MS models would make the UA computationally intractable, as one run of such a model may take hours [41]. Two approaches have been discussed in the literature for reducing the computational time of UA in MS models: “emulator based” [31, 41, 45] and “sample size based” [39].

Emulator-based approaches approximate the solution space of the simulation model by preassumed functions mapping the input parameters to the outcome. These approximate functions predict the outcome of the main simulation model instead of running the time-consuming PMS at each instance of the MC process [31, 41, 45]. The types of approximate functions used in MS disease models include linear response function used by Cronin et al. [31] for a model of prostate cancer, with the objective of investigating the effect of screening on prostate cancer mortality rates, and Gaussian function for a patient-level model of osteoporosis in Stevenson et al. [41].

The emulator-based method would save a significant amount of computation time at a cost of losing precision due to outcome approximation. Stevenson et al. [41] reported that the required time for a single run of their model was reduced from 150 minutes to “virtually instantaneous”. However, this method requires the analyst to assume a certain distribution for the outcome in order to be able to fit a response surface. For example, Cronin et al. [31] used a linear approximate function and assumed Poisson distribution for the number of prostate cancer-related deaths, which was the outcome of choice for performing UA in their study. Complex PMS models would need a specific response function for each outcome and there is no *a priori* reason to expect that all response functions have the same functional form. In other words, one needs to use a different approximate function for each type of outcome in a single model.

The sample size approach involves finding a combination of values for the simulated population size and number

of simulation runs that achieve a specified precision for the outcome within a fixed computational time [39, 45]. O'Hagan et al. [39] used analysis of variance (ANOVA) to determine sample size in the case where two types of uncertainty, MC error and parameter uncertainty, are considered. Although their approach was originally designed for decision analytic models with incremental net benefit as the outcome, it can be used for other outcomes, including those typically observed in PMS models. O'Hagan et al. [39] reported that the computational demand was reduced by a factor of almost 36 (They mentioned that this claim is for the incremental net benefit outcome and the resulting efficiency gain could be different for other outcomes.). The sample size approach is simple to implement and does not require an approximation function.

UA in MS models has been discussed in detail for cost-effectiveness studies with decision analytic purposes that use specific type of outcome (i.e., cost-effectiveness ratio or incremental net benefit) [47]. However, systematic presentations of UA for predictive PMS models with multiple outcomes are lacking. In particular, there has been little discussion of UA in the context of chronic disease models with epidemiological or burden-of-disease outcomes, such as those predicting the incidence and prevalence of a disease in a population, its impact on mortality and quality of life, or its costs [3, 12]. In the following section we provide an overview of the key steps in UA for such models.

3. Steps for Performing Uncertainty Analysis

3.1. Selecting the Outcomes. UA should be performed separately for each outcome of interest. Due to their predictive nature, most of PMS models are dynamic; that is, the outcomes are time dependent. Typical examples of such dynamic outcomes in public health applications include the distribution of health determinants, disease incidence or prevalence, mortality, quality of life, health care utilization, and costs. The selection of the outcomes for UA depends on the objectives of the study.

3.2. Selecting Parameters. After selecting the outcome of interest (e.g., the prevalence of disease X over the next N years), the next step is to screen the parameters and select the relevant ones [45]. By "relevant", we mean those parameters that are likely to have significant influence over the selected outcome, as determined by the model analyst. It is important to note that the uncertainty intervals from UA will be conditional on the assumption that all parameters not selected for UA do not introduce additional uncertainty. Although ideally one would include all model parameters in the UA, this is often impractical in complex models, such as many PMS disease models, that may include hundreds of parameters [1, 2, 5, 7].

Tools for determining parameter inclusion range from qualitative approaches, such as scrutinizing the model's conceptual diagram and investigating its major components, to more quantitative approaches, such as parameter prioritizing [45] and tornado diagrams [37, 47]. Scrutinizing the model, beginning with its major components, helps in identifying

the most important parameters, eliminating the irrelevant ones, and identifying possible dependencies between the parameters.

Different components of the simulation model, such as demographic, disease, and intervention components, need to be examined to ascertain which parameters to include in the UA. The data sources for the parameters should be considered as well. For example, the demographic component may include the parameters describing the distribution of age and sex in the population, as well as birth and death rates over time. These parameters are often derived from vital statistics or large population databases and may be excluded from UA due to minimal uncertainty. However, if any of the above processes are modeled from a sample, then sampling variability should be incorporated into the UA. For example, a CISNET model for breast cancer [16] used Surveillance, Epidemiology, and End Results (SEER) [51] to estimate all-cause and cancer-specific mortality rates by age and sex. Because the SEER database is very large, the sampling variability of the SEER-derived mortality rates is likely small, but may not be negligible. Further analyses, discussed later in this section, may be needed to determine the inclusion of such parameters in the UA.

In making decisions about which parameters to include, the analyst needs to take into account possible discrepancies between the population being simulated and the base population from which the data have been gathered. If the calibration has been performed to fill in this discrepancy, the parameters used in the calibration should be included in the UA (unless they are estimated from a large population data or census data). In addition, the calibration algorithm itself should be integrated into the UA process. For example, in POHEM-OA [46], a PMS model of osteoarthritis in Canada, the baseline incidence of osteoarthritis (OA) was estimated from a population-based administrative database in British Columbia (BC), Canada, that includes healthcare data for almost all residents of the province (Table 1) [52]. To simulate the incidence of OA in Canada, a calibration algorithm is needed to obtain the incidence of OA in the reference category of BMI, given the BMI hazard ratios and the distribution of BMI in the population. The calibration algorithm uses a direct search method to match the marginal distribution of incidence in the province of BC to that of the simulated population of the Canada [52] by changing the baseline distribution until the target value is reached for every age/sex category within a small acceptable distance. We will discuss the details of implementing calibration into UA in Section 3.4.

Parameters in disease components of the PMS models are mostly coefficients in statistical models used to determine the distribution of risk factors in the population, both the baseline distribution and the dynamics of how risk factors change through time and space, in addition to the coefficients used in modeling the probability of event, mostly through relative risks (or hazard ratios) associated with the risk factors. For example, in POHEM-OA [46], the baseline distribution of body mass index (BMI), a risk factor for OA, was derived from the Canadian Community Health Survey (CCHS) [50] (Table 1). The dynamics of how BMI

TABLE 1: Components and parameters of a simplified version of POHEM-OA* that was used for the uncertainty analysis example.

Demographic components	
	Age distribution in 2001**
	Sex distribution in 2001**
	Province of residence distribution in 2001**
	Birth and mortality rates by age and sex over time#
	Migration##
Disease components	
(i)	Baseline parameters
	BMI [‡] distribution in 2001**
	Incidence of OA in 2001 by sex and 5-year age groups ^{‡‡}
	Reference (“baseline”) hazard rates of OA [§]
	Prevalence of OA in 2001 by sex and 5-year age groups ^{§§}
(ii)	Risk factors model
	Change in BMI over time ^{§*}
(iii)	Disease incidence
	Effect of BMI on incident OA by sex ^{§*}

* POHEM-OA: Population Health Microsimulation model for osteoarthritis (OA) [46]. ** Observed in Canadian Community Health Survey (CCHS-2001) [50]. # Projected mortality in Canada by age and sex from Statistic Canada. ## Migration data obtained from Statistics Canada projections. [‡] BMI: weight/height². ^{‡‡} From administrative data in BC [52]; OA is defined as at least 2 visits to a health professional within 2 years or 1 hospitalization with the ICD-9 code 715. Incident cases in 2001 are identified after excluding prevalent cases prior to 2001 using a 10 year run-in period [46]. [§] Obtained numerically using calibration to match the marginal distribution of incidence in BC administrative data. ^{§§} Obtained as the final stable prevalence from a simulation of the Canadian population over a 50-year horizon, under constant age-specific incidence rates. ^{§*} Obtained from a linear regression model including age, sex, province, education, and prior BMI. ^{§#} BMI was categorized into 4 standard categories (see Table 3). The effect (hazard ratio) for each level of BMI is obtained from a survival regression model using longitudinal NPHS data (two cycles: 2000 and 2002) [49], separately for men and women.

changes in time and in different provinces was modeled from the longitudinal National Population Health Survey (NPHS) [49]. Parameters in disease component can be excluded from UA if they are obtained from full population data, such as the census or vital statistics. However, this is often not the case for the parameters used in modeling probability of events (e.g., relative risks associated with risk factors) as they are mostly estimated from sample data [27]. For example, the coefficients in Framingham equations, commonly used for modeling the risk of cardiovascular disease in simulation models, come from the Framingham sample [53] and, therefore, are subject to sampling variation.

Another group of parameters that should be considered for UA are those describing the effects of health interventions [54]. For example, in the chronic disease prevention (CDP) model [55], a PMS model for evaluating interventions for reducing obesity, the effects of physician counseling was translated into reductions in the percentage of total cholesterol intake, BMI, and systolic blood pressure. Another example is the parameters describing the effects of a screening program on quality adjusted life years (QALYs) and excess mortality in a CISNET model of breast cancer [16].

Whenever the outcome chosen for the UA (in Section 3.1) is related to the intervention, the best practice is to include the “intervention” parameters into UA, as most of these parameters are estimated from sample data and the uncertainty associated with them significantly affect the outcome uncertainty.

The tornado diagram is a useful tool for prioritizing the parameters in the initial steps of UA [47]. A tornado diagram represents the results of multiple univariate sensitivity analyses in which one parameter is changed at a time and the highest and lowest value of the outcome are determined. The parameters are then sorted according to their impact and presented graphically, often as a horizontal bar diagram [25]. Tornado diagrams do not consider the interactions and correlations between parameters. Figure 2 shows an example of the tornado diagram for the POHEM-OA model discussed in greater detail later in Section 4.

3.3. Assigning Distributions to Parameters (or Using Bootstraps). For the selected parameters, the model analyst should either assign a distribution or use a bootstrapping approach to be able to apply the MC method and get the uncertainty interval for the outcome, as discussed in Sections 3.4 and 3.5. Parameters in the simulation models may define the risk of an event, such as hazards ratios, relative risks, odds ratios, and transition probabilities of events, or estimate the dynamic changes in a continuous variable (e.g., regression coefficients for modelling BMI changes over time [46]). Parameters may be estimated directly from data, drawn from literature, or based on expert opinions. For parameters estimated from data using parametric (and semiparametric) statistical models, the distribution can be obtained from the relevant statistics. The common regression models, such as the parametric multiple linear, logistic, and semiparametric Cox proportional hazards models, utilize a specific distribution for the outcome [47].

Incorporating uncertainty into the risk of an event depends on the structure of the model. In some predictive MS models, risk functions for events use event-history modeling [56, 57] as opposed to memoryless (Markovian) states [9, 47, 58]. Event-history modeling equations represent the behaviour of an individual through dependence on his/her simulated past. For example, in the LifePaths model [59], a PMS model of the Canadian labour force [60], the timing of a child’s birth is determined by the timing of parents’ marriage, the time since the previous births in the family, and the age of the parents. In continuous-time event-history models with hazard functions, an asymptotic lognormal distribution can be assigned to hazard ratios [47], while a normal (Gaussian) distribution can be assigned to the log transformation of the relative risk or odds ratio, under the large sample assumption [61].

Regression coefficients used in modelling changes of risk factors over time are another types of parameter in PMS models. For instance, coefficients in the dynamic models of BMI changes in POHEM include age, sex, place of birth (time-invariant covariates), and socio-economic status (time-dependent covariate), as well as previous levels of BMI [46]. For such complex models, the uncertainty analyst needs

to estimate the variance of the coefficients of the regression models (or use bootstrap weights for the outcome) [60].

In MS models utilizing longitudinal data, events are often defined based on hazard functions and waiting time distributions [61, 62]. The waiting time distribution can be exponential (for fixed hazards) [46], Weibull (for time-dependent hazards) [47, 61], or Gompertz (for time-dependent hazards with additional complexity, such as those for tumour growth modeling in cancer models [16]). Transition probabilities can be converted into hazards for use in continuous-time models [5, 62]. Other examples of parameter distributions often used in health simulation models include the beta distribution (e.g., for modeling cost data)[55], the gamma distribution (e.g., for modeling quality-adjusted life years) [47], and the Poisson distribution for modeling event counts [31].

For parameters drawn from the literature, only the mean and its 95% confidence interval are often reported, without the actual distribution. In such cases, method of moments [47] can be used. Alternatively, the analyst needs to assume a specific distribution appropriate for a given parameter. The variance of the parameter can then be estimated from the confidence interval reported by the authors [47].

Parameters estimated by expert opinion are often ignored in UA. This may result in a substantial underestimation of the true uncertainty [27]. It is often possible to assign upper and lower bounds to the estimated parameters using experts' judgment, for example, where there are biological limits to a disease process (e.g., tumour growth) or limits to an epidemiological parameter such as prevalence or case fatality. In these cases, the distribution could be quantified by a uniform, triangular, or trapezoidal distribution by assigning lower and upper bounds in addition to the mode of the distribution, or lower and upper bounds on the mode in case of a trapezoidal distribution [63]. General distributions such as the generalized beta distribution family and the Johnson translation system of distributions are among the best choices in case several data points have been selected in the literature or gathered from expert opinions [63]. Another approach to assigning distribution to parameters estimated from expert opinion is the Bayesian method [64]. A key task in applying this method is the expression of expert opinion in the form of a (prior) probability distribution [64]. An example is the Sheffield elicitation framework (SHELF) developed by Garthwaite et al. [64], elaborating steps to construct a probability distribution for the parameters expressed by a group of experts.

For parameters estimated from nonparametric models, the uncertainty analyst should use bootstrap methods instead of sampling from a distribution. Bootstrap method provides an easier approach for estimating uncertainty for those parameters whose distribution cannot be identified. It can also be used to reduce the computational difficulties in sampling from a parametric distribution for parameters estimated from surveys that use a complex sampling design involving stratification and clustering [65]. In both cases, bootstrap methods sample with replacement from the original data [66]. As an example, the National Population Health Survey (NPHS) [49], a longitudinal household survey

conducted by Statistics Canada every two years since 1994, provides bootstrap weights for users to estimate the variance of their estimators, such as regression coefficients. We will implement this type of bootstrapping in Section 4.

Accounting for Correlations between Parameters. After selecting the parameters for UA, the analyst needs to examine the relationship between the parameters to ensure their correlations are taken into account. The analyst can identify the correlations between parameters by examining the model diagram and reviewing the parameter estimation process [25]. For example, relative risks associated with different levels of an ordinal risk factor are often correlated. As illustrated in Figure 1, in the POHEM-OA model, the hazard ratios of OA incidence for four different BMI categories are interdependent. Such dependence should be reflected in this step of the UA, through assigning a joint distribution using covariance matrices (often estimated from the regression model). When bootstrap methods are used, parameter interdependence will be inherently reflected in the distribution of the outcome.

3.4. Monte Carlo Method. This step involves applying the MC method to calculate the uncertainty intervals for the mean outcome of interest. This step involves sampling from the assigned distribution or using bootstrap methods as discussed in Section 3.3.

Sampling Approach. For those parameters that have been assigned a distribution, we need to specify the sampling approach, as it will affect the uncertainty intervals of the outcome. The sampling approach should be both effective, to encompass the entire distribution of the parameter, and efficient, to sample only the required number of points. There are several sampling approaches for the MC method [45]. One of the most effective is the Latin hypercube sampling (LHS) [31, 43]. LHS is a particular type of stratified sampling, which samples from the equally probable segments of the parameters distribution. It is designed to provide a comprehensive coverage of the parameter space. More complex sampling approaches include orthogonal sampling and adaptive kriging sampling [44, 45]. Although LHS is not the most efficient sampling approach, as it samples more points than necessary, it is commonly used due to its effectiveness. It samples enough points within a reasonable computational time and is more efficient than random uniform sampling.

A common sampling technique is to generate a uniform random number (between 0 and 1) through one of the existing random number generator algorithms [61] (e.g., pseudorandom number generators are among the most used methods that deterministically perform series of tasks to generate statistically sound (uniform) random numbers [12]). Then, using the inverse of the cumulative density function (CDF), known as the quantile function, one finds a real value of the parameter.

For correlated parameters, an appropriate joint distribution should be specified, as discussed in Section 3.3, and

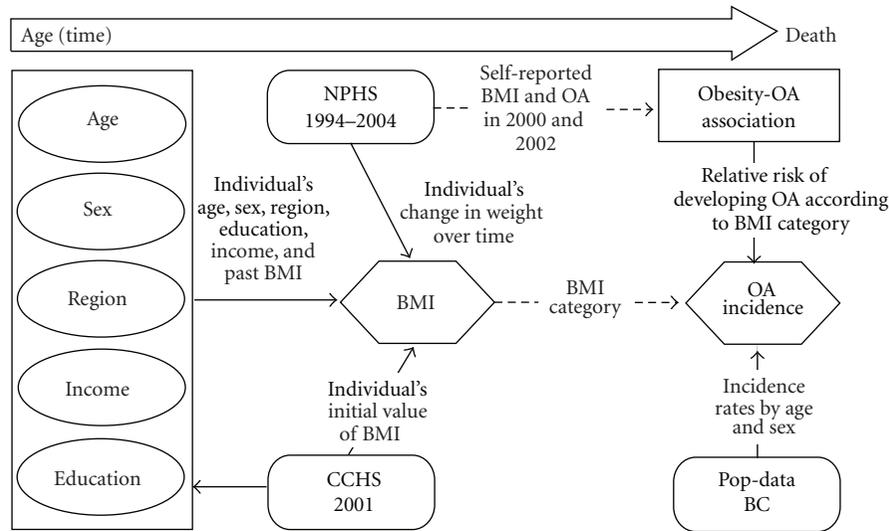


FIGURE 1: Diagram of the simplified version of the POHEM-OA used for the uncertainty analysis example. POHEM-OA: Population Health Microsimulation model for osteoarthritis (OA) [46]. Rectangles inside the figure are data sets used for estimation of parameters; first in top left: NPHS 1994–2004: National Population Health Survey, a longitudinal household survey by statistics Canada with a sample of size of over 17,000 persons that started in 1994 [49], 5 cycles (1996–2006) have been used in BMI trajectory model and two cycles (2000–2002) used in estimation of relative risks (hazard ratios) of OA. **Rectangle in the bottom left corner: CCHS 2001 (Canadian Community Health Survey [50]) used for generating the starting population for the microsimulation with characteristics listed on the left column of the figure. **Rectangle in the bottom right corner: Pop-data BC (population data BC [49]): administrative billing data for the province of BC including data for >98% residents of BC from 1994–2003, used for estimating incidence of OA.

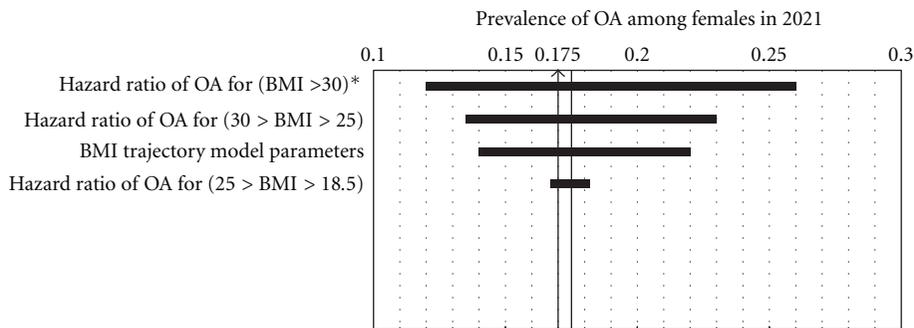


FIGURE 2: Tornado diagram for prevalence of osteoarthritis among females in year 2021—predicted by POHEM-OA. POHEM-OA: Population Health Microsimulation model for osteoarthritis (OA) [46]; *each of the hazard ratios parameters on the left have been varied based on the upper and lower limits of the estimated 95% CI shown in Table 3.

the sampling technique should be adapted accordingly [47]. In general, if the marginal univariate distributions of the parameters and the correlation between them are known, sampling from a multivariate distribution can be performed using the multivariate normal copula with matrix algebra (e.g., Cholesky decomposition) [47].

Calibration in UA. Calibration is a technique commonly used in population-based models with predictive goals when direct estimation of a given parameter is not possible and model parameters are selected so that the model reproduces observed results [67]. At each step of the MC sampling, calibration should be performed with the new (random) sample of the parameters, generated in previous step. After the calibration procedure converges, the simulation model

should be run with the calibrated parameter. As a result, the calibration procedure should be implemented inside the UA and be executed at each run of the MC method.

Determining the Number of MC Runs and Population Size. The computational burden of UA can be reduced by the appropriate choice of the simulated population size (m) and number of simulation runs (n). As discussed in O’Hagan et al. [39], the variance of the mean outcome based on the simulation results is given by the formula:

$$\hat{\sigma}^2 = \left(\frac{\hat{\sigma}_1^2}{mn} \right) + \frac{\hat{\sigma}_2^2}{n}. \tag{1}$$

In (1) above, $\hat{\sigma}_1^2$ reflects the MC error and $\hat{\sigma}_2^2$ reflects the parameter uncertainty (and it is assumed that patients

outcomes are independent). The goal of the sample size approach is to find values of (m) and (n) such that a desired precision level d is achieved; that is, ($\hat{\sigma}^2 = d$) is achieved given a fixed amount of computation time. The computational time is assumed to be linear in terms of both population size (m) and number of MC runs (n) (i.e., $K = mn$, where K is measured in terms of number of individuals that can be simulated in the available time.). Since $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ are unknown, the solution is obtained in two steps. In the initial step, a small number of simulation runs are completed and $\hat{\sigma}^2$ and $\hat{\sigma}_1^2$ are estimated from the outcome while $\hat{\sigma}_2^2$ is obtained by subtraction using (1). Next, in the main step, (1) is solved for (n) with (mn) replaced by K and $\hat{\sigma}^2$ replaced by (d) to obtain the lowest number of computational runs, (n^*), that satisfies the computational time and precision constraints. The corresponding value for the population size, (m^*), is then obtained from the constraint $K = m^*n^*$. Note that this procedure typically does not generate the optimal choices for (m^*) and (n^*). The final estimated number of MC runs is the smallest number that can produce the desired precision level for a given computational time.

As discussed in O'Hagan et al. [39], in order to gain the maximum information for estimating uncertainty intervals in a fixed amount of computational time, one should use a relatively small population size and a large number of simulation runs. In other words, the optimal design is to simulate a small population over and over again using a large number of sample points from the joint distribution of the parameters.

3.5. Constructing the Empirical Distribution of the Outcome. In the final step of the UA, the analyst constructs the cumulative distribution function (CDF) of the outcome after applying the MC method with a population size of (m) and number of MC runs of (n) calculated using the "sample size" approach discussed in the previous section. The 95% uncertainty interval of the outcome is constructed using the quantile function based on the CDF of the outcome being analyzed. The results of the UA, often represented by 95% uncertainty intervals, should be presented as conditional on the assumptions of the model.

4. Example of Uncertainty Analysis in POHEM-OA

We provide an example of the UA approach outlined in the previous section, using a simplified version of POHEM-OA, a PMS model of osteoarthritis in Canada (Figure 1) [46]. POHEM-OA is a complex model, featuring a large number of parameters [46]. In our example, the focus is on modeling the incidence and prevalence of OA. Table 1 lists the parameters used in this example and the data source for each. Table 2 describes the steps of the proposed UA approach and the corresponding steps of the POHEM-OA example. In the first step of the proposed UA (Section 3.1), we selected the gender-specific prevalence of OA in Canada from 2001 to 2021 as the outcome of interest. In the second step of the UA (Section 3.2), we first investigate the conceptual model diagram shown in Figure 1. Most of the

parameters shown in Figure 1 are not included in the UA. Parameters estimated from the Canadian Community Health Survey (CCHS-2001) [50], a national survey with a very large sample size ($N = 130,000$), and parameters obtained from provincial administrative data in BC [52] are excluded because their variance is considered to be small.

For the selected parameters, a Tornado diagram has been constructed as shown in Figure 2. Two types of parameters have shown the highest impact on uncertainty in univariate analysis and were included in the final UA analysis. These are (i) the hazard ratios (HR's) for the effect of different levels of BMI on OA incidence and (ii) the regression coefficients used in modeling BMI progression in the simulated individuals over time. The HR's for BMI categories have been estimated from the National Population Health Survey (NPHS) [49], a longitudinal survey in Canada with a sample of size of over 17,000 persons that started in 1994. Two cycles of NPHS were used for the analysis of HR's in POHEM-OA. The HR's have been estimated using an exponential survival regression model [46]. The point estimates and 95% confidence intervals for the parameters are shown in Table 3 and the correlation matrices are shown in Table 4. The model for BMI progression is based on 6 cycles of the NPHS (1996–2006) [49] and includes BMI history (up to 4 past BMI values) and other covariates such as age, sex, region, education, and income. The model was stratified by age, sex, and BMI, resulting in 112 regression models (28 age-sex-BMI strata and 4 autoregression models).

In the third step of the proposed UA (Section 3.3), we used a multivariate lognormal distribution for the hazard ratios. In the fourth step of the proposed UA (Section 3.4), Latin hypercube sampling (LHS) has been used to sample from the multivariate HR's distribution with the given correlation matrices (Table 4), translated into covariance matrices for the LHS. The second sets of parameters are those used in the BMI progression models. To avoid the complexity of assigning distributions to a large number of models, we used the bootstrap weights developed for the NPHS by Statistics Canada [49]. For each BMI progression model (age and sex specific), we derived eight sets of parameter estimates, meaning that each set of parameters represents a sample from the underlying joint multivariate distribution for the covariates of the BMI model. At each run of the UA, we randomly selected one of these samples based on each person's age/sex, in addition to randomly sampling the hazard ratio of the OA event from a multivariate lognormal distribution. POHEM-OA calibrates the baseline incidence of OA using population data from the province of British Columbia [52]. As discussed in the calibration step of Section 3.4, we have automated the calibration algorithm and integrated it with the MC method, so that for each sample of HR's, calibration is performed to get the (calibrated) baseline incidence rate; then, repetitive runs of the simulation model are performed by applying the MC method.

To determine the number of MC runs and the population size (last step of Section 3.4), we used the method discussed by O'Hagan et al. [39]. The number of MC runs is calculated as a function of precision level (set to 0.01), computational

TABLE 2: Proposed steps for uncertainty analysis in population-based microsimulation models and corresponding steps for POHEM-OA uncertainty analysis example.

Steps for performing uncertainty analysis in population-based microsimulation models	Steps for uncertainty analysis in POHEM-OA
Step 1. Selecting the outcome	Sex-specific prevalence in (2001–2021)
Step 2. Establishing the UA list of parameters	(1) Hazard ratios for each BMI-category (2) BMI progression
Step 3. Assigning a probability density function to parameters (or use bootstrap sampling)	(1) Lognormal distribution for hazard ratios with sex-specific correlation matrices (2) Using 8 alternative regression models using bootstrap weights from NPHS ^s (1996–2006)
Step 4. Applying Monte Carlo method	(1) Latin hypercube sampling (2) Random sampling with replacement (among eight alternate sets of parameters)
(i) Select a sampling approach (or use bootstrap sampling).	Calibrate on incidence by age and sex; implement the calibration algorithm into the MC method, using squared error criteria for convergence.
(ii) Implement calibration into MC method.	Result: $m_0 = 20,000,000$ population size; $n_0 = 5$ (for the initial run); $m^* = 500,000$ people; $n^* = 785$ MC-runs for $d = 0.01$; for 12-hour run of a PC with 12 GB memory, CPU = i7-980 Intel, 3.3 GHz.
(iii) Calculate n^* , m^* for precision level d .	
Step 5. Constructing the outcome distribution	Please see Figure 3.

^s NPHS: National Population Health Survey, a longitudinal household survey by statistics Canada with a sample of size of over 17,000 persons that started in 1994 [49].

TABLE 3: Point estimates and 95% CI for hazard ratios of osteoarthritis in POHEM-OA*, estimated from the NPHS** (2000–2002); reference category: normal weight ($18.5 \leq \text{BMI} < 25$).

	Male (95% CI)	Female (95% CI)
Under weight ($\text{BMI} < 18.5$)	—	0.33 (0.02–0.4)
Normal weight ($18.5 \leq \text{BMI} < 25$)	1 (ref)	1 (ref)
Over weight ($25 \leq \text{BMI} < 30$)	1.07 (0.68, 1.74)	1.76 (1.14–2.62)
Obese ($\text{BMI} \geq 30.0$)	1.69 (1.03, 2.81)	2.03 (1.34, 2.96)

* POHEM-OA: Population Health Microsimulation model for osteoarthritis (OA) [46]; **NPHS: National Population Health Survey, a longitudinal household survey by statistics Canada with a sample of size of over 17,000 persons that started in 1994 [49].

time (set to 12 hours), and initial estimates of the MC error and parameter uncertainty. The initial estimates were obtained by running the model 5 times with a sample size of 20 million (Each POHEM run that simulates the entire Canadian adult population of around 20 million takes about 35 minutes on a PC with 12 GB RAM and 3.33 GHz. CPU, Intel i7-X980.). Applying the formulas from O’Hagan et al. [39], as described in Section 3.4, we obtained ($n = 758$) and ($m = 500,000$). Approximately 10–15 iterations were needed for calibration before each main run, and each run with population size of ($m = 500,000$) took about 50 seconds. It should be noted that without applying the sample size approach by O’Hagan et al. [39], to get the same precision

level with a sample size of 20 million, we would need to run the model about 400 times and it would take about 12 days of run time (an almost 25-fold increase in time efficiency). Although increasing the population size (m) or increasing the number of MC runs (n) increases the precision level of the mean estimate (and consequently the precision of the resultant uncertainty interval), increasing (n) has a much higher impact on the precision level than increasing (m). The reason is that the variance of the mean estimator ((1) in Section 3.4) is a function of (m) and (n), where (n) affects both terms on the right-hand side of the equation, and (m) only affects one term, as discussed in O’Hagan et al. [38]. That is why running the model 400 times with a sample of 20 million produces the same precision as running the model 785 times with a sample of 500,000.

Result of the UA is shown in Figure 3. The estimated 95% uncertainty intervals for the prevalence of OA in Canada in 2021 are 0.09 to 0.18 for men and 0.15 to 0.23 for women if we include both MC error and parameter uncertainty of the hazard ratios of OA associated with different categories of BMI as well as BMI trajectory model parameters. These results are based on the assumption that uncertainty in other parameters shown in Table 1 can be ignored. This assumption would be reasonable for parameter estimated from large population-based datasets, such as provincial administrative data [52], but not for parameters estimated from smaller studies, as discussed in Section 3.2. The uncertainty contributed by the parameters estimated from the CCHS (sample size 130,000) can be assumed to be small, compared to those estimated from the NPHS [49] (sample size of 17,000), and has been ignored in this

TABLE 4: Correlation matrix for hazard ratios of osteoarthritis in POHEM-OA*, estimated from the NPHS **(2000–2002); reference category: normal weight ($18.5 \leq \text{BMI} < 25$), shown for male (female) in the upper (lower) triangular part of the matrix.

	BMI < 18.5	$18.5 \leq \text{BMI} < 25$	$25 \leq \text{BMI} < 30$	BMI ≥ 30.0
BMI < 18.5	1	0.5125	0.5362	(Ref)
$18.5 \leq \text{BMI} < 25$	0.6234	1	0.5405	—
$25 \leq \text{BMI} < 30$	0.6325	0.5812	1	—
BMI ≥ 30.0	(Ref)	—	—	—

* POHEM-OA: Population Health Microsimulation model for osteoarthritis (OA) [46]; **NPHS: National Population Health Survey, a longitudinal household survey by statistics Canada with a sample of size of over 17,000 persons that started in 1994 [49].

example. If the uncertainty associated with other parameters were included in the analysis, the UA interval would have been equal or larger.

5. Discussion

Complex MS models are increasingly used in modeling chronic diseases [1, 4–9, 17–21]. UA is critical to the validation of such models and interpretation of their results [24–30]. There is a growing literature on the methods of capturing uncertainty in quantitative policy analyses [30, 68]. The US Environmental Protection Agency [40] and the National Academy of Science [32] identify MC methods as preferred approaches to quantifying uncertainty in environmental risk assessment models.

The goal of UA in MS models is to provide uncertainty intervals around each outcome. Applying the MC method to estimate the variance of the mean outcome is the core of the UA process. Several guidelines discuss the steps in performing UA to capture the uncertainty in risk assessment models [32] and cost-effectiveness models in health economics [34]. However, there is a gap in the literature regarding the application of MC methods to UA in predictive PMS models. To our knowledge, no UA guidelines have been developed specifically for MS models that incorporate health outcomes other than cost-effectiveness ratios and simulate real populations rather than hypothetical cohorts.

In this paper we suggest an approach for estimating the uncertainty intervals for PMS models that implicitly takes into account the complexity of the model, number of parameters, model calibration, and correlations among the parameters. We have outlined the steps involved in performing UA, including a practical approach to lowering the computational demand, developed by O’Hagan et al. [39]. The key idea that led to the sample size approach of O’Hagan et al. [39] is the trade-off between the number of MC runs and the population size. In fact, by increasing the number of MC runs while decreasing the population size, one can increase the precision level of the resultant uncertainty interval for a fixed computational time [39]. This result has been discussed in both Rutter et al. [13] and Griffin et al. [29] for the optimal design of the UA. The major assumption is that the computational time is linear in m and n .

In contrast to the optimal design, UA in MS models is often performed by running the model a small number of times, due to the computational time constraint [13, 29].

However, this practice is not acceptable due to the low precision of the resultant intervals. On the other hand, ignoring first-order uncertainty (i.e., MC error) would result in a biased, and too small, uncertainty estimate of the outcome in MS models. The approach applied in this paper can be used to reduce the computation time while taking into account both the first- and the second-order uncertainty. Alternatively, one can use response surface approaches such as linear regression surfaces [31] and Gaussian process modeling [41] to reduce the required number of runs. The response surface methods can be also combined with the sample size approach applied in this paper from O’Hagan et al. [39]. Future studies can compare these approaches in terms of their performances in estimating uncertainty intervals for MS models within a fixed computational time. It is worth emphasizing that the results of UA should be interpreted as conditional on the assumptions involved in performing this analysis. For example, ignoring the structural uncertainty should be reported in conjunction with the resulting uncertainty intervals. In addition, the sample size (population size and the number of runs) and the precision level used in sample size calculations should be reported.

In the example of UA provided in Section 4, we have implemented the proposed approach for POHEM-OA. The resultant 95% uncertainty intervals incorporate the uncertainty associated with the hazard ratios for the effect of BMI on OA risk and use bootstraps to include the BMI trajectory model parameters. By adopting the sample size approach developed by O’Hagan et al. [39], we were able to reduce the computational time from 12 days to 12 hours, with the same level of precision for the uncertainty intervals.

The UA approach described in this study is not free of limitations. One limitation is that methods specific to agent-based models with interindividual interactions were not addressed. However, our approach can be generalized to include such models in future studies. Another limitation is that we only addressed two sources of uncertainty, MC error and parameter uncertainty. Other sources, specifically, uncertainty due to model structure and sources of data, have been ignored. Draper [69] discussed the Bayesian approach for structural uncertainty and demonstrated how alternative statistical model structures, such as functional forms for dose-response relationships or alternative link function for generalized linear models, can be included in the UA. Berry et al. [16] performed this type of structural UA for breast cancer models developed by CISNET. CISNET consists of seven model development groups who work independently but

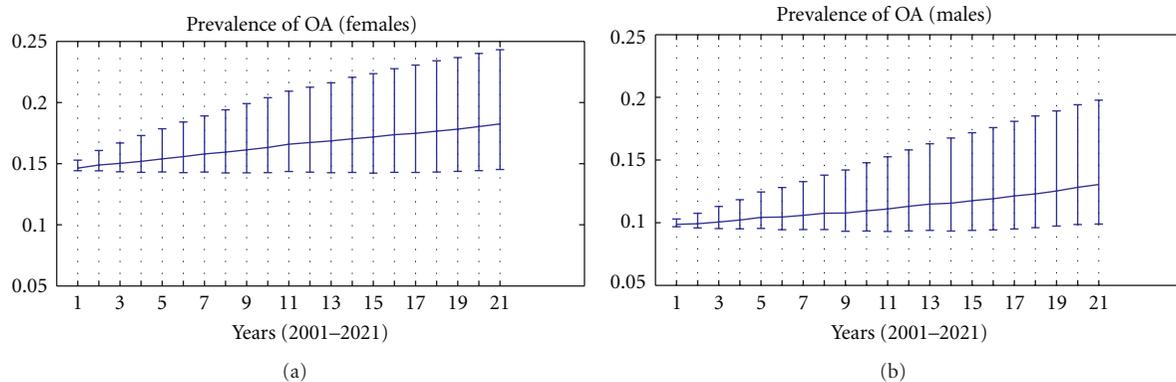


FIGURE 3: Results of uncertainty analysis: sex-specific prevalence of osteoarthritis (OA) in Canada during (2001–2021) as predicted by POHEM-OA; curves inside the graphs represent the mean estimate surrounded by 95% uncertainty intervals. POHEM-OA: Population Health Microsimulation model for osteoarthritis (OA) [46].

with the same standardized calibration data and protocols. The authors regarded outcomes from the seven models to be a sample from a larger population of possible model results and used kernel density estimation, a nonparametric approach, to estimate the probability density function of the outcome that includes model uncertainty [16]. However, individual models in CISNET that used MS framework were not subjected to systematic UA. As noted in Rutter et al. [13] for MS models with health outcomes and Poulter et al. [32] for environmental risk assessment models, performing quantitative UA for structural uncertainty is feasible only when all model developers use the same data input and cooperate to simulate and present the outcome. Unless such cooperation between model developers can be assured, structural uncertainty should be approached in the model validation process through scenario analyses, rather than being incorporated into the quantitative UA [32].

We have approached the UA from the traditional statistical viewpoint, the frequentist statistics, which begins with the assumption that the factors to be estimated are unknown constants. In contrast, Bayesian inferences begin with the assumption that every parameter is stochastic [42, 69]. As such, they directly incorporate uncertainty of the outcome in a simulation model. Bayesian methods incorporating Markov Chain Monte Carlo (MCMC) are increasingly used to model correlated parameters in UA instead of bootstrap-based methods or parametric distributions discussed here [68, 69].

Finally, it should be noted that UA is a prerequisite for “value of information analysis” in decision-analytic models [33, 47] or stochastic “sensitivity analysis (SA)” as defined in environmental models [32], where contribution of each parameter to the total uncertainty of the model is being estimated [45, 70]. SA and UA are warranted when a decision must be made about whether to expand resources to acquire additional information with regard to input parameters of the simulation model. In general, the greater the uncertainty of the outcome, the greater the expected values of additional information gained by future research [47]. UA can be used to estimate the sensitivity indices of each parameter [45] to determine the value of collecting additional data in order

to improve the accuracy of the model’s prediction [33, 70]. Future studies should be directed toward providing methods of stochastic SA in PMS models.

Acknowledgments

The authors would like to thank Nick Bansback, with Ph.D. degree from University of British Columbia, Vancouver, BC, Canada, Saeideh Dahaghin with a Ph.D. degree from Arthritis Research Centre of Canada, Vancouver, BC, Canada, Claude Nadeau with a Ph.D. degree from Statistics Canada, Ottawa, ON, Canada, William M. Flanagan, with a B.M. degree from Statistics Canada, Ottawa, ON, Canada, Keiko Asakawa with a Ph.D. degree from Statistics Canada, Ottawa, ON, Canada and the STAR team: Douglas G. Manuel, MD, with M.S. degrees from Statistics Canada, Ottawa, ON, Canada and University of Ottawa, Ottawa, ON, Canada, David Buckeridge, MD, with Ph.D. degree from McGill University, Montreal, QC, Canada, Jillian Oderkirk, with M.S. degree from Statistics Canada, Ottawa, ON, Canada, Michal Abrahamowicz, with Ph.D. degree from McGill University, Montreal, QC, Canada, Sam Harper, with Ph.D. degree from McGill University, Montreal, QC, Canada, Anya Okhmatovskaia, with M.S. degree from McGill University, Montreal, QC, Canada, and Mushfiqur Rahman, with M.S. degrees from University of British Columbia, Vancouver, BC, Canada and Arthritis Research Centre of Canada, Vancouver, BC, Canada, and Canadian Arthritis Network for partially providing funding for Behnam Sharif.

References

- [1] S. K. Lhachimi, W. J. Nusselder, H. C. Boshuizen, and J. P. Mackenbach, “Standard tool for quantification in health impact assessment. A review,” *American Journal of Preventive Medicine*, vol. 38, no. 1, pp. 78–84, 2010.
- [2] M. C. Weinstein, “Recent developments in decision-analytic modelling for economic evaluation,” *Pharmacoeconomics*, vol. 24, no. 11, pp. 1043–1053, 2006.
- [3] E. Ackerman, “Simulation of micropopulations in epidemiology: tutorial 1—simulation: an introduction,” *International*

- Journal of Bio-Medical Computing*, vol. 36, no. 3, pp. 229–238, 1994.
- [4] E. J. Feuer, R. Etzioni, K. A. Cronin, and A. Mariotto, “The use of modeling to understand the impact of screening on US mortality: examples from mammography and PSA testing,” *Statistical Methods in Medical Research*, vol. 13, no. 6, pp. 421–442, 2004.
 - [5] B. P. Will, J. M. Berthelot, K. M. Nobrega, W. Flanagan, and W. K. Evans, “Canada’s Population Health Model (POHEM): a tool for performing economic evaluations of cancer control interventions,” *European Journal of Cancer*, vol. 37, no. 14, pp. 1797–1804, 2001.
 - [6] J. N. Struijs, M. L. L. van Genugten, S. M. A. A. Evers, A. J. H. A. Ament, C. A. Baan, and G. A. M. van den Bos, “Modeling the future burden of stroke in the Netherlands: impact of aging, smoking, and hypertension,” *Stroke*, vol. 36, no. 8, pp. 1648–1655, 2005.
 - [7] J. Mandelblatt, C. B. Schechter, W. Lawrence, B. Yi, and J. Cullen, “The SPECTRUM population model of the impact of screening and treatment on U.S. breast cancer trends from 1975 to 2000: principles and practice of the model methods,” *Journal of the National Cancer Institute Monographs*, no. 36, pp. 47–55, 2006.
 - [8] M. C. Wolfson, “POHEM—a framework for understanding and modelling the health of human populations,” *World Health Statistics Quarterly*, vol. 47, no. 3–4, pp. 157–176, 1994.
 - [9] A. Brennan, S. E. Chick, and R. Davies, “A taxonomy of model structures for economic evaluation of health technologies,” *Health Economics*, vol. 15, no. 12, pp. 1295–1310, 2006.
 - [10] B. Nosyk, B. Sharif, H. Sun, C. Cooper, and A. H. Anis, “The cost-effectiveness and value of Information of three influenza vaccination dosing strategies for individuals with human immunodeficiency virus,” *PLoS ONE*, vol. 6, no. 12, Article ID e27059002705, 2011.
 - [11] S. Zhao, Z. Xu, and Y. Lu, “A mathematical model of hepatitis B virus transmission and its application for vaccination strategy in China,” *International Journal of Epidemiology*, vol. 29, no. 4, pp. 744–752, 2000.
 - [12] D. Wolf, “The role of microsimulation in longitudinal data analysis. Papers in microsimulation series,” *Canadian Studies in Population*, vol. 28, no. 2, pp. 313–339, 2001.
 - [13] C. M. Rutter, A. M. Zaslavsky, and E. J. Feuer, “Dynamic microsimulation models for health outcomes: a review,” *Medical Decision Making*, vol. 31, no. 1, pp. 10–18, 2011.
 - [14] N. Anders Klevmarken, “Statistical inference in micro-simulation models: incorporating external information,” *Mathematics and Computers in Simulation*, vol. 59, no. 1–3, pp. 255–265, 2002.
 - [15] R. T. Hoogenveen, P. H. M. van Baal, H. C. Boshuizen, and T. L. Feenstra, “Dynamic effects of smoking cessation on disease incidence, mortality and quality of life: the role of time since cessation,” *Cost Effectiveness and Resource Allocation*, vol. 6, article 1, 2008.
 - [16] D. A. Berry, K. A. Cronin, S. K. Plevritis et al., “Effect of screening and adjuvant therapy on mortality from breast cancer,” *The New England Journal of Medicine*, vol. 353, no. 17, pp. 1784–1792, 2005.
 - [17] H. J. Smolen, D. J. Cohen, G. P. Samsa et al., “Development, validation, and application of a microsimulation model to predict stroke and mortality in medically managed asymptomatic patients with significant carotid artery stenosis,” *Value in Health*, vol. 10, no. 6, pp. 489–497, 2007.
 - [18] M. Hoogendoorn, M. P. M. H. Rutten-van Mólken, R. T. Hoogenveen et al., “A dynamic population model of disease progression in COPD,” *European Respiratory Journal*, vol. 26, no. 2, pp. 223–233, 2005.
 - [19] F. Loeve, R. Boer, G. J. van Oortmarssen, M. van Ballegooijen, and J. D. F. Habbema, “The MISCAN-COLON simulation model for the evaluation of colorectal cancer screening,” *Computers and Biomedical Research*, vol. 32, no. 1, pp. 13–33, 1999.
 - [20] A. J. Palmer, S. Roze, W. J. Valentine et al., “Validation of the CORE diabetes model against epidemiological and clinical studies,” *Current Medical Research and Opinion*, vol. 20, supplement 1, pp. S27–S40, 2004.
 - [21] R. Kahn, P. Alperin, D. Eddy et al., “Age at initiation and frequency of screening to detect type 2 diabetes: a cost-effectiveness analysis,” *The Lancet*, vol. 375, no. 9723, pp. 1365–1374, 2010.
 - [22] S. F. Railsback, S. L. Lytinen, and S. K. Jackson, “Agent-based simulation platforms: review and development recommendations,” *Simulation*, vol. 82, no. 9, pp. 609–623, 2006.
 - [23] D. Perrin, H. J. Ruskin, and M. Crane, “Model refinement through high-performance computing: an agent-based HIV example,” *Immune Research*, vol. 6, supplement 1, article S3, 2010.
 - [24] R. G. Sargent, “Verification and validation of simulation models,” in *Proceedings of the 37th Winter Simulation Conference (WSC ’05)*, pp. 130–143, December 2005.
 - [25] J. A. Kopec, P. Finès, D. G. Manuel et al., “Validation of population-based disease simulation models: a review of concepts and methods,” *BMC Public Health*, vol. 10, article 710, 2010.
 - [26] C. F. Citro and E. A. Hanushek, *Improving Information for Social Policy Decisions: The Uses of Microsimulation Modeling*, National Research Council, National Academy Press, New York, NY, USA, 1991.
 - [27] J. A. Salomon, C. D. Mathers, C. J. L. Murray, and B. Ferguson, “Methods for life expectancy and healthy life expectancy uncertainty analysis,” GPE Discussion Paper, World Health Organization, Geneva, Switzerland, August 2001, <http://www.who.int/healthinfo/paper10.pdf>.
 - [28] A. H. Briggs, A. E. Ades, and M. J. Price, “Probabilistic sensitivity analysis for decision trees with multiple branches: use of the dirichlet distribution in a Bayesian framework,” *Medical Decision Making*, vol. 23, no. 4, pp. 341–350, 2003.
 - [29] S. Griffin, K. Claxton, N. Hawkins, and M. Sculpher, “Probabilistic analysis and computationally expensive models: necessary and required?” *Value in Health*, vol. 9, no. 4, pp. 244–252, 2006.
 - [30] M. G. Morgan and M. Henrion, *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*, Cambridge University Press, Cambridge, UK, 1992.
 - [31] K. A. Cronin, J. M. Legler, and R. D. Etzioni, “Assessing uncertainty in microsimulation modelling with application to cancer screening interventions,” *Statistics in Medicine*, vol. 17, no. 21, pp. 2509–2523, 1998.
 - [32] S. R. Poulter, “Monte Carlo simulation in environmental risk assessment—science, policy and legal Issues,” *Risk: Health, Safety & Environment*, vol. 7, pp. 7–26, 1998.
 - [33] G. Baio and A. P. Dawid, “Probabilistic sensitivity analysis in health economics,” *Statistical Methods in Medical Research*, <http://www.ucl.ac.uk/statistics/research/pdfs/rr292.pdf>. In press.
 - [34] K. Claxton, M. Sculpher, C. McCabe et al., “Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra,” *Health Economics*, vol. 14, no. 4, pp. 339–347, 2005.

- [35] J. Hay and J. Jackson, "Panel 2: methodological issues in conducting pharmaco-economic evaluations—modeling studies," *Value in Health*, vol. 2, no. 2, pp. 78–81, 1999.
- [36] B. G. Koerkamp, M. C. Weinstein, T. Stijnen, M. H. Heijnenbroek-Kal, and M. G. M. Hunink, "Uncertainty and patient heterogeneity in medical decision models," *Medical Decision Making*, vol. 30, no. 2, pp. 194–205, 2010.
- [37] M. C. Weinstein, B. O'Brien, J. Hornberger et al., "Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR task force on good research practices—modeling studies," *Value in Health*, vol. 6, no. 1, pp. 9–17, 2003.
- [38] P. Doubilet, C. B. Begg, M. C. Weinstein, P. Brun, and B. McNeil, "Probabilistic sensitivity analysis using Monte Carlo simulation. A practical approach," *Medical Decision Making*, vol. 5, no. 2, pp. 157–177, 1985.
- [39] A. O'Hagan, M. Stevenson, and J. Madan, "Monte Carlo probabilistic sensitivity analysis for patient level simulation models: efficient estimation of mean and variance using ANOVA," *Health Economics*, vol. 16, no. 10, pp. 1009–1023, 2007.
- [40] D. E. Burmaster and P. D. Anderson, "Principles of good practice for the use of Monte Carlo techniques in human health and ecological risk assessments," *Risk Analysis*, vol. 14, no. 4, pp. 477–481, 1994.
- [41] M. D. Stevenson, J. Oakley, and J. B. Chilcott, "Gaussian process modeling in conjunction with individual patient simulation modeling: a case study describing the calculation of cost-effectiveness ratios for the treatment of established osteoporosis," *Medical Decision Making*, vol. 24, no. 1, pp. 89–100, 2004.
- [42] J. E. Oakley and A. O'Hagan, "Probabilistic sensitivity analysis of complex models: a Bayesian approach," *Journal of the Royal Statistical Society Series B*, vol. 66, no. 3, pp. 751–769, 2004.
- [43] B. P. Will, K. M. Nobrega, J. M. Berthelot et al., "First do no harm: extending the debate on the provision of preventive tamoxifen," *British Journal of Cancer*, vol. 85, no. 9, pp. 1280–1288, 2001.
- [44] M. D. McKay, R. J. Beckman, and W. J. Conover, "Comparison of three methods for selecting values of input variables in the analysis of output from a computer code," *Technometrics*, vol. 21, no. 2, pp. 239–245, 1979.
- [45] A. Saltelli, M. Ratto, T. Andres et al., *Global Sensitivity Analysis: The Primer*, John Wiley & Sons, Chichester, UK, 2008.
- [46] J. A. Kopec, E. C. Sayre, W. M. Flanagan et al., "Development of a population-based microsimulation model of osteoarthritis in Canada," *Osteoarthritis and Cartilage*, vol. 18, no. 3, pp. 303–311, 2010.
- [47] A. Briggs, K. Claxton, and M. Sculpher, *Decision Modeling for Health Economic Evaluation. Handbooks in Health Economic Evaluation*, Oxford University Press, London, UK, 2006.
- [48] C. M. Rutter, D. L. Miglioretti, and J. E. Savarino, "Bayesian calibration of microsimulation models," *Journal of the American Statistical Association*, vol. 104, no. 488, pp. 1338–1350, 2009.
- [49] Statistics Canada, "National population health survey data (NPHS)," Survey information, 2011, <http://www23.statcan.gc.ca:81/imdb/p2SV.pl?Function=getSurvey&SDDS=3225&lang=en&db=imdb&adm=8&dis=2>.
- [50] Statistics Canada, "Canadian community health survey (CCHS) cycle 1.1," Public use micro-data file, 2011, <http://www.statcan.gc.ca/cgi-bin/imdb/p2SV.pl?Function=getSurvey&SurvId=3226&SurvVer=0&InstaId=15282&InstaVer=1&SDDS=3226&lang=en&db=imdb&adm=8&dis=2>.
- [51] National Cancer Institute, "Surveillance epidemiology and end results (SEER)," US National Institute of Health, 2011, <http://seer.cancer.gov/popdata/download.html>.
- [52] "Population Data BC, (British Columbia, Canada)-Multi-university platform," British Columbia's pan-provincial population health data services, 2011, <http://www.popdata.bc.ca>.
- [53] A. I. Adler, "UKPDS—modelling of cardiovascular risk assessment and lifetime simulation of outcomes," *Diabetic Medicine*, vol. 25, supplement 2, pp. 41–46, 2008.
- [54] S. K. Lhachimi, W. J. Nusselder, P. van Baal et al., "DYNAMO-HIA: a dynamic microsimulation model-model specification for a dynamic model for health impact assessment," User Guide and Manual-3rd EUPHA Conference, Amsterdam, The Netherlands, 2010, [http://www.dynamo-hia.eu/object_binary/o2925_Simulation Brief.pdf](http://www.dynamo-hia.eu/object_binary/o2925_Simulation%20Brief.pdf).
- [55] F. Sassi, M. Cecchini, J. Lauer, and D. Chisholm, "Improving lifestyles, tackling obesity: the health and economic impact of prevention strategies," OECD Health Working Papers 48, OECD, 2009, <http://dx.doi.org/10.1787/2200087432153>.
- [56] H. Blossfeld and G. Rohwer, *Techniques of Event History Modelling: New Approaches to Causal Analysis*, Lawrence Erlbaum Associates, New Jersey, NJ, USA, 2nd edition, 2002.
- [57] D. T. Levy, P. L. Mabry, Y. C. Wang et al., "Simulation models of obesity: a review of the literature and implications for research and policy," *Obesity Reviews*, vol. 12, no. 5, pp. 378–394, 2011.
- [58] M. Hilgsmann, O. Ethgen, O. Bruyère, F. Richy, H. J. Gathon, and J. Y. Reginster, "Development and validation of a markov microsimulation model for the economic evaluation of treatments in osteoporosis," *Value in Health*, vol. 12, no. 5, pp. 687–696, 2009.
- [59] Statistics Canada, "The lifepaths microsimulation model (version 1.1): an overview. Information of the model," 2011, http://www.statcan.gc.ca/sp/sd/LifePathsOverview_E.pdf.
- [60] G. Rowe and H. Nguyen, "Longitudinal analysis of labour force survey data," *Survey Methodology (Statistics Canada)*, vol. 30, no. 1, pp. 105–114, 2004.
- [61] F. Willekens, "Continuous-time microsimulation in longitudinal analysis," in *New Frontiers in Microsimulation Modelling*, A. Zaidi, A. Harding, and P. Williamson, Eds., pp. 413–436, Ashgate, Surrey, UK, 2009.
- [62] R. Bender, T. Augustin, and M. Blettner, "Generating survival times to simulate Cox proportional hazards models," *Statistics in Medicine*, vol. 24, no. 11, pp. 1713–1723, 2005.
- [63] M. E. Kuhl, J. S. Ivy, E. K. Lada, N. M. Steiger, M. A. Wagner, and J. R. Wilson, "Univariate input models for stochastic simulation," *Journal of Simulation*, vol. 4, no. 2, pp. 81–97, 2010.
- [64] P. H. Garthwaite, J. B. Kadane, and A. O'Hagan, "Statistical methods for eliciting probability distributions," *Journal of the American Statistical Association*, vol. 100, no. 470, pp. 680–701, 2005.
- [65] D. J. Pasta, J. L. Taylor, and J. M. Henning, "Probabilistic sensitivity analysis incorporating the bootstrap: an example comparing treatments for the eradication of *Helicobacter pylori*," *Medical Decision Making*, vol. 19, no. 3, pp. 353–363, 1999.
- [66] D. Yeo, H. Mantel, and T. P. Liu, "Bootstrap variance estimation for the National Population Health Survey," *Proceedings of Survey Research Methods Section, The American Statistical Association*, vol. 12, no. 1, pp. 778–785, 1999.
- [67] R. Dowling, A. Skabardonis, J. Halkias, G. McHale, and G. Zammit, "Guidelines for calibration of microsimulation models: framework and applications," *Transportation Research Record*, vol. 1876, no. 1, pp. 1–9, 2004.

- [68] G. Parmigiani, "Measuring uncertainty in complex decision analysis models," *Statistical Methods in Medical Research*, vol. 11, no. 6, pp. 513–537, 2002.
- [69] D. Draper, "Assessment and propagation of model uncertainty," *Journal of the Royal Statistical Society Series B*, vol. 57, pp. 45–97, 1995.
- [70] A. M. Zaslavsky and S. W. Thurston, "Error analysis of food stamp microsimulation models: further results," in *Proceedings of the ASA Social Statistics Section*, pp. 151–156, American Statistical Association, Alexandria, Va, USA, 1995.



Hindawi
Submit your manuscripts at
<http://www.hindawi.com>

