Comparing simulation run time requirements to achieve stabilized absolute and incremental costs effectiveness results in probabilistic vs. deterministic modeling analysis

Authors: Volker Foos [1], Phil McEwan [2], and David Grant [3]

1 IMS Health, Basel, Switzerland; 2 Centre for Health Economics, Swansea University, United Kingdom; 3 IMS Health, London, United Kingdom;

Introduction

The predominant rationale for parameter sampling in decision analytic modeling is to assess parameter uncertainty, i.e. to determine the uncertainty of the modeling outcomes given the uncertainty of the input parameters. This is commonly evaluated in the context of probabilistic sensitivity analysis (PSA).

It is less conclusive if parameters should be sampled in standard base case analysis (BCA) where the intention is to achieve mean end point (EP) predictions close enough to expected value (EV). The justification for sampling in BCA is to capture patient heterogeneity but also account for nonlinear effects (systematic deviation of mean EP predictions in sampled vs. non sampled projections induced by the nonlinear model characteristics).

However recent HTA recommendations suggested conducting BCA without parameter sampling for the reason that EP stabilization may be negatively influenced and correspondingly run time requirements (RTR) increased.

In stochastic modeling approaches that utilize Monte Carlo techniques the random variability of outcomes (Monte Carlo Error) can be reduced through run time increases (increase of patient number processed through the modeling).

Objectives

The objective of our investigation was to explore the extent to which parameter sampling distorts endpoint stabilization in stochastic modeling approaches.

We further aimed to assess the degree of mean EP prediction failure in non sampled (deterministic) analyses (where patient heterogeneity and nonlinear effects are not considered) versus sampled (probabilistic) analyses.

Methods

This study used the IMS Core Diabetes Model (CDM) (1-3), a validated and established diabetes model, to compare the EP stabilization patterns of analyses with and without parameter sampling.

Model projections were obtained evaluating the cost effectiveness of two hypothetical interventions for patients with characteristics as presented in Table 1 and differences in clinical effectiveness of 0.5% HbA1c in favor of the treatment- vs. control arm.

In order to understand the degree by which input parameter variability distorts EP stabilization, model projections were run in three ways:

A) Deterministic: no parameter sampling applied
B) Low degree of sampling (L_SMP): parameters sampling with standard error (SE) ~1% of mean
C) High degree of sampling (H_SMP): parameters were sampled with SE ~5% of mean

Further, sampling analyses were conducted in two alternative ways:

1st to include all possible sources of uncertainty (i.e. patient baseline characteristics, treatment effects, probabilities, costs and utilities)
2nd to omit sampling of economic parameters (costs and utilities) since their contribution to constitute patient heterogeneity and nonlinear effects is limited.

All costs, life expectancy (LE) and quality adjusted life expectancy (QALE) were projected over a life time horizon (50 years) in non-parametric bootstrap simulations with 1000 included patients repetitively run through a maximum of 10,000 bootstrap iterations.

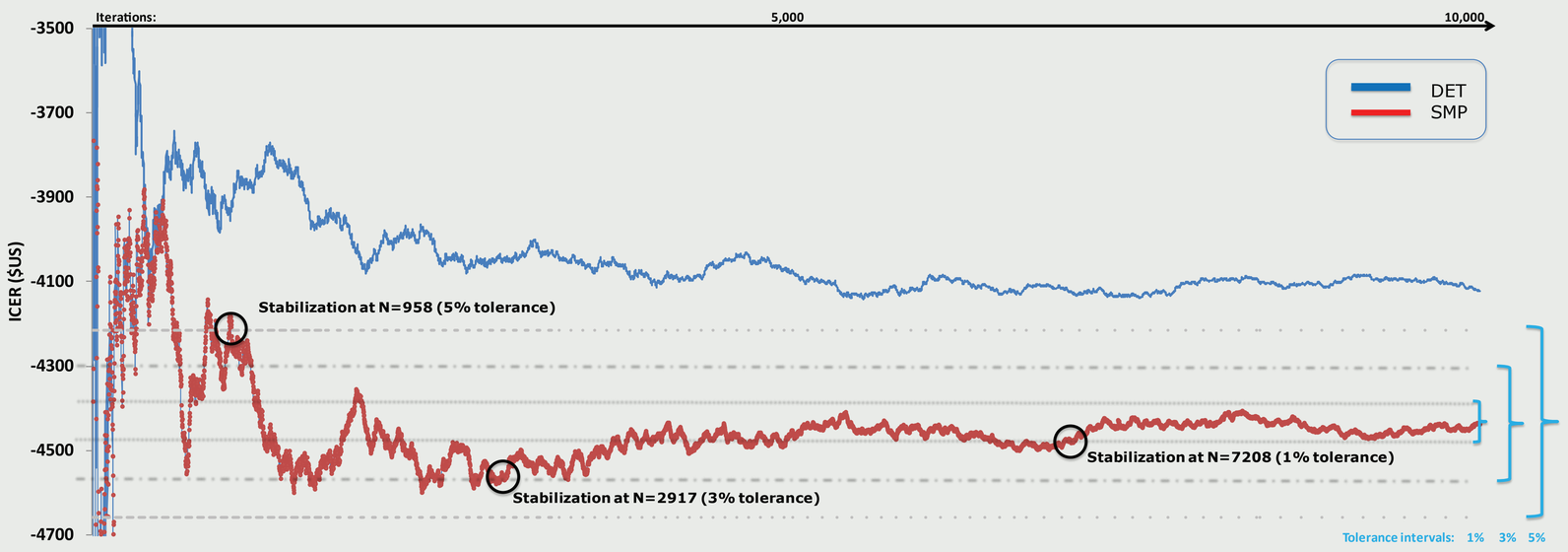
Costs (US$) and benefits were discounted at 3.0%.

RTR were assessed in terms of bootstrap replications required to reach stabilization of absolute and incremental EPs.

Stabilization was defined as point estimates remaining within the interval of expected value (EV) +/- tolerance (%); tolerance was explored in a range of 0.1% to 5% surrounding EV (Figure 1). EV was assumed to be equal to the EP outcome at 10,000 iterations.

Last, to evaluate the impact on non-incorporation of patient heterogeneity and related non-linear effects, we assessed the % degree of bias in deterministic mean EP prediction vs. EV. The latter was assumed to correspond to mean probabilistic outcome at maximum run time (10,000 bootstrap iterations).

Figure 1) Mean ICER (based on LE) stabilization over 10,000 bootstrap iterations with tolerance levels from 1%, 3% and 5%



DET = deterministic (non-sampled; SMP = sampled

Figure 2) EP stabilization in deterministic analysis

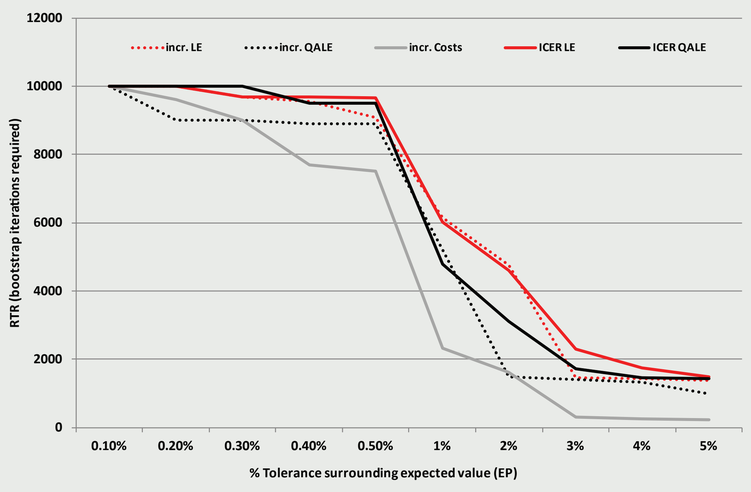


Figure 3) EP stabilization in probabilistic analysis (SE=1% of mean)

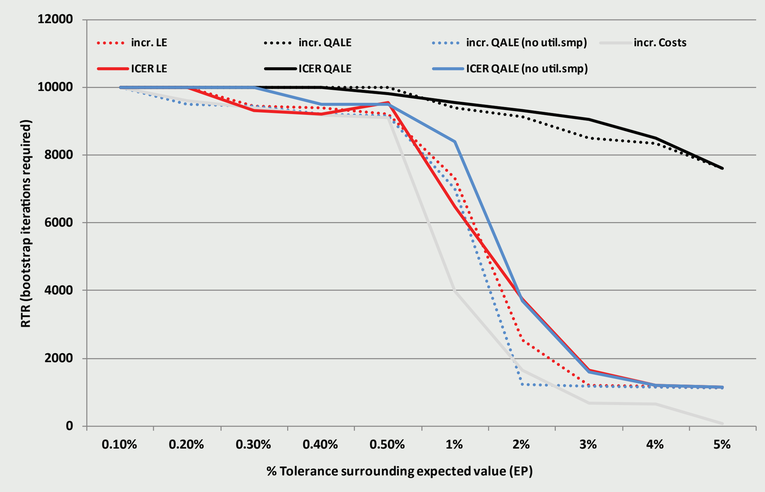


Table 1) Patient characteristics

	(M+S)	(M+D)	(BI)
Current age	55	0.473	2.25
Gender	Male		
Ethnicity	White		
Duration since diagnosis of T2DM	5	0.198	0.25
(years)	7.49	0.046	0.375
% HbA1c	200	2.059	10
Total cholesterol (mg/dl)*	100	1.482	5
LDL cholesterol (mg/dl)*	47	0.482	2.35
HDL cholesterol (mg/dl)*	265	3	13.25
Systolic blood pressure (mmHg)	133.6	0.692	12.65
BMI (kg/m2)	30	0.194	1.5
Baseline complication history	0%	0%	0%

Figure 4) EP stabilization in probabilistic analysis (SE=5% of mean)

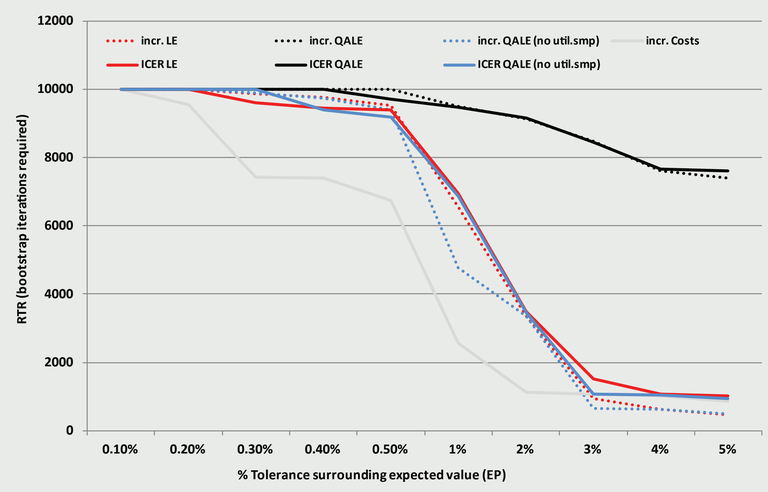
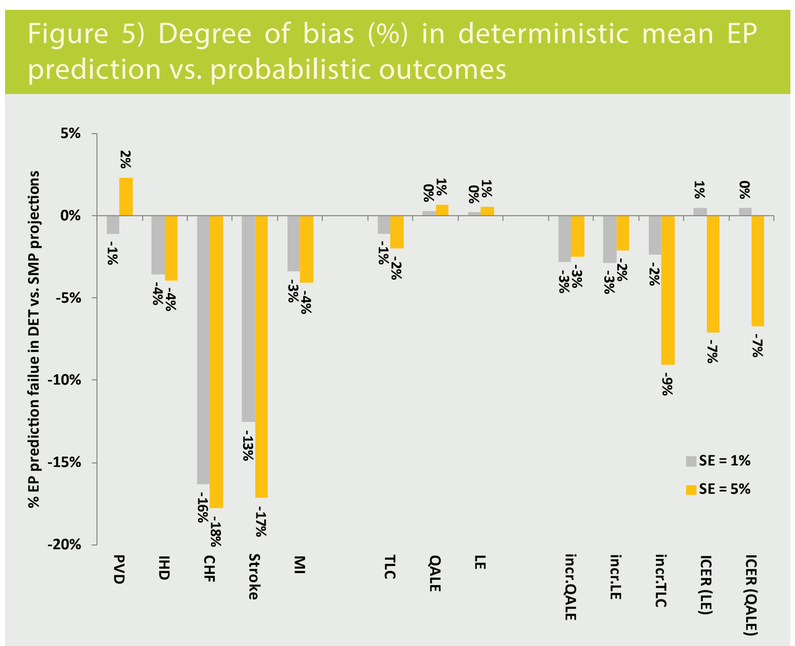


Figure 5) Degree of bias (%) in deterministic mean EP prediction vs. probabilistic outcomes



Results

Stabilization of absolute EP including LE, QALE, total lifetime costs and complication incidence occurred with replications <1000 in both, SMP and DET analyses.

RTR to reach stabilized incremental results were considerably greater with ICERs (per LE or per QALE) consistently exposing the lowest stabilization characteristics in both, deterministic and probabilistic analyses (Figures 2 to 4).

In deterministic analysis 6020, 4770, 2300, 1750 and 1500 replications were required to reach stabilization for tolerance intervals ranging from 1% to 5%, respectively (Figure 2).

Probabilistic analyses exposed considerably greater RTR if economic parameters (utilities and costs) were subjected to sampling and the number of bootstrap iterations required to reach stabilization determined by the largest tolerance interval of 5% surrounding EV was > 7,000 (Figures 3 and 4).

In probabilistic analyses where the sampling of economic parameters was omitted, RTR were comparable to those achieved in deterministic analysis with 8400, 3750, 1650, 1217 and 1156 replications required to reach stabilization for tolerance intervals ranging from 1% to 5% using 1% SE (Figure 3) and 6954, 3475, 1522, 1074 and 1034 replications required using 5% SE (Figure 4).

The % degree of bias in deterministic mean EP prediction vs. probabilistic outcomes (assumed to correspond to EV) at maximum run time of 10,000 bootstrap iterations are presented in Figure 5.

Conclusions

The consideration of parameter sampling in standard BCA does not add additional RTR if economic parameters such as costs and utilities are omitted from sampling.

EP stabilization appears to be not influenced by the degree of parameter sampling (size of confidence ranges in which parameters are sampled).

In contrast, failure to accommodate patient heterogeneity and associated nonlinear effects within the modeling can significantly bias predicted morbidity, morality and cost effectiveness findings.

When treatment decision rules are dependent on heterogeneous patient attributes, probabilistic sampling of associated parameters should be routinely undertaken in standard BCA.

References

[1] Palmer et al. Curr Med Res Opin 2004;20:S5–S26.
[2] Palmer et al. Curr Med Res Opin 2004;20:S27–40.
[3] Foos, V et al. Value in Health 2012;15(7)p.A470

FOR FURTHER INFORMATION: Please contact Phil McEwan (phil.mcewan@heor.co.uk) or Volker Foos (vfoos@ch.imshealth.com)
IMS HEALTH | 210 PENTONVILLE ROAD, LONDON N1 9JY, UNITED KINGDOM