Original Article

How to Address Uncertainty in Health Economic Discrete-Event Simulation Models: An Illustration for Chronic Obstructive Pulmonary Disease

Isaac Corro Ramos (iD)**, Martine Hoogendoorn, and Maureen P. M. H. Rutten-van Mölken**

Background. Evaluation of personalized treatment options requires health economic models that include multiple patient characteristics. Patient-level discrete-event simulation (DES) models are deemed appropriate because of their ability to simulate a variety of characteristics and treatment pathways. However, DES models are scarce in the literature, and details about their methods are often missing. **Methods**. We describe 4 challenges associated with modeling heterogeneity and structural, stochastic, and parameter uncertainty that can be encountered during the development of DES models. We explain why these are important and how to correctly implement them. To illustrate the impact of the modeling choices discussed, we use (results of) a model for chronic obstructive pulmonary disease (COPD) as a case study. **Results**. The results from the case study showed that, under a correct implementation of the uncertainty in the model, a hypothetical intervention can be deemed as cost-effective. The consequences of incorrect modeling uncertainty included an increase in the incremental cost-effectiveness ratio ranging from 50% to almost a factor of 14, an extended life expectancy of approximately 1.4 years, and an enormously increased uncertainty around the model outcomes. Thus, modeling uncertainty incorrectly can have substantial implications for decision making. **Conclusions**. This article provides guidance on the implementation of uncertainty in DES models and improves the transparency of reporting uncertainty methods. The COPD case study illustrates the issues described in the article and helps understanding them better. The model R code shows how the uncertainty was implemented. For readers not familiar with R, the model's pseudo-code can be used to understand how the model works. By doing this, we can help other developers, who are likely to face similar challenges to those described here.

Keywords
patient-level model, discrete event simulation model, uncertainty, heterogeneity, COPD, personalized medicine

Key Points

In health economic (HE) decision modeling, the lack of reporting extensive details on the model implementation, especially on modeling uncertainty, is often a great concern. Since the majority of the HE models published so far, including those using patient-level data, are basically Markov models, this issue is of special importance for discrete-event simulation (DES) models.

Our article can be used as an example on how to appropriately implement uncertainty in DES models and transparently report the methods used. This can be useful to other model developers, who are likely to face similar challenges to those described in our article. We have used a chronic obstructive pulmonary disease DES model as a case study to help the reader to get a better understanding of the problems presented in our article. By including

Corresponding Author
Isaac Corro Ramos, Institute for Medical Technology Assessment, Erasmus University Rotterdam, P.O. Box 1738, Rotterdam, Zuid-Holland, 3000 DR, The Netherlands. (corroramos@imta.eur.nl).

the full R code of the COPD model, readers can see how the proposed solutions were implemented. For readers who are not familiar with R, we have provided the model's pseudo-code, which can be used as a standalone tool to understand how the flow of the code works without knowing the specifics of the R language.

Background

Pharmacological and nonpharmacological treatments are increasingly targeted to the patients who are most likely to benefit. Personalized medicine, which includes stratified medicine, refers to an approach where treatments are targeted to subgroups of patients with specific characteristics and takes into account other morbidities frequently occurring simultaneously with the condition at hand.[1] The economic evaluation of these personalized treatment options calls for innovations in economic modeling.

Health economic (HE) models built to assess the cost-effectiveness of new treatments in stratified medicine should be flexible enough to include many patient and disease characteristics that are deemed important for disease prognosis or treatment allocation. Such models can be used to calculate a range of different outcomes and to evaluate treatment options for a wide variety of subgroups. Markov models may not be the most efficient way to address the heterogeneity in patient, disease, and treatment characteristics that need to be considered to target the right treatment to the right patient. Most of the HE models published so far, including those using patient-level data, are basically Markov models.[2–6] Patient-level discrete-event simulation (DES) models are appropriate tools to model treatments in stratified medicine because of their ability to simulate a greater variety of (time-varying) patient characteristics and treatment pathways.[7–9] DES models also allow a more flexible time management, where time to multiple (and possibly competing) events is simulated, without restricting their occurrences to predetermined model cycles. DES models can also accommodate the patient history into the simulation and do not require the definition of health states.[10–12] Despite this, DES models have not been widely used for economic evaluations.[13] The methods

used in DES models are usually complex and their description rather technical. When DES models are used in published economic evaluations, details about their implementation are often not part of the main article or even the supplementary materials. Multiple examples of published studies using patient-level models not reporting extensive details about modeling uncertainty can be found in the literature.[14–31] This lack of guidance is not limited to DES models.[32] However, since DES models are less frequently used and are often considered more complex, DES models could benefit more from additional modeling guidance. Considering also the increased attention for personalized medicine, it is argued that DES models will become more important because they are more flexible in modeling individual patient's disease and treatment pathways based on many different characteristics.[33]

Existing modeling guidelines focus on what should be covered by DES models but are not very specific on how.[8] The Technical Support Document 15 (TSD15) by the Decision Support Unit (DSU) of the National Institute for Health and Care Excellence (NICE) provides more details about the methodology used in DES models,[7] including the code of a simple DES model used to perform cost-effectiveness analysis. As models become more complex, new methodological challenges may arise, which are not covered by TSD15. This specifically pertains to uncertainty in DES models. When methods to model uncertainty are not implemented or explained well, the credibility of the DES model's results may decrease. The aim of this article is to describe important challenges of modeling heterogeneity and structural, stochastic, and parameter uncertainty in DES models. We describe the solutions to several issues faced during the development of a new DES model for chronic obstructive pulmonary disease (COPD) and present the results of hypothetical cost-effectiveness analyses to illustrate the impact of each type of uncertainty on the model outcomes.[34] These are common issues that can be encountered in other DES models. Thus, the methods explained in this article could be applied to other DES models as well. This article is further structured into a Methods section, in which 4 challenges and solutions to modeling uncertainty in DES models are described; a Results section, in which the impact of the applied solutions is illustrated; and a Discussion section.

Methods

Description of the COPD Model

A full description of the HE COPD model was published in Hoogendoorn et al.[34] The main objective of the model

Institute for Medical Technology Assessment (iMTA), Erasmus University Rotterdam, Rotterdam, Zuid-Holland, The Netherlands (ICR, MH, MPRM); Erasmus School of Health Policy & Management (ESHPM), Erasmus University Rotterdam, Rotterdam, The Netherlands (MPRM).

was to simulate major COPD-related events over a lifetime and to calculate HE outcomes such as total costs and quality-adjusted life years (QALYs) for a population of COPD patients. The backbone of the model is a set of regression equations describing the associations between patient and disease characteristics and different COPD-related events, as well as intermediate and final COPD outcomes. These equations were estimated using the combined patient-level data of 5 large COPD trials with 1 to 4 years of follow-up after randomization (19,378 patients).[35–39] The regression equations are outlined in Table 1. A summary of the characteristics of the trial populations is shown in Appendix 1 (available online).

The simulation starts by sampling patients (with replacement) from the existing baseline populations. The model combines these regression equations in the sequence of events described in Figure 1. Model inputs are individual patient characteristics. The events included in the model were exacerbations, pneumonias, and death. Final outcomes were total lifetime costs and QALYs. A step-by-step description of the simulation is also given in Figure 1.

The simulation base-case represents patients treated with a common medication for COPD patients (tiotropium). The effectiveness of new interventions is modeled relative to the base-case and can be included in the model by modifying inputs of the base-case arm (e.g., 15% increase in the time to exacerbation compared to the base-case).

Challenges of Modeling Uncertainty during the Model Implementation

Simulating patients' clinical histories is the core of the COPD model. The full R code used to build the COPD model can be found on GitHub (https://github.com/icor-roramos/PLDES_COPD_model). Future updates in the code will be placed on the same repository. We refer to the model pseudo-code in several parts of the remaining of this article, with the purpose of improving clarity and transparency. The pseudo-code for the clinical history function is shown in Figure 2. The full model pseudo-code is presented in Appendix 2 (available online). The COPD model includes the 4 types of uncertainty that are relevant for decision models: patient heterogeneity and stochastic, parameter, and structural uncertainty.[44] Challenges associated with each type of uncertainty that were encountered during the model implementation are described below.

Challenge 1: Remove differences in patient heterogeneity between the intervention and control arms

Why is this challenge important? Patient heterogeneity is defined as the variability in model outcomes between patients that can be attributed to differences in patient characteristics.[44] Because patient-level models usually sample individual patients, whose characteristics are used as inputs for regression equations, model results are influenced by patient heterogeneity. Sampling different patients for the intervention and control groups may cause differences in the model results that are attributed to differences in heterogeneity rather than to differences in effectiveness between treatments.

What solution was implemented? In the COPD model, patient heterogeneity is the result of randomly sampling different patients at the start of the simulation. Patients are sampled with replacement from the baseline populations, and every time a simulation is run, results could be different when a different set of patients is sampled. The way to solve this is by using random seeds, which are numbers that, after being *fixed* in the code, ensure that the random draws in the model can be reproduced. This would produce the same model outcomes in all simulations run with the same seed. This is relevant when 2 treatment arms are modeled: the population has to be the same for both arms to get results that differ only because of a treatment effect but not due to a different selection of patients. To avoid this issue in the COPD model, the random seed 1 in the clinical history function shown in Figure 2 has to be the same for all the treatment arms compared.

Challenge 2: Adjusting remaining life expectancy after the occurrence of an event

Why is this challenge important? Structural uncertainty is the type of uncertainty associated with the assumptions made while building a model. It may include very different types of assumptions, from the choice of the software used to estimate regression equations, to the way remaining life expectancy is adjusted after events occur in the simulation. This is important because the choices made during the development of a model have to some extent impact on the model results.

What solution was implemented? The way remaining life expectancy was adjusted after events occur in the COPD model is used here for illustrative purposes. As explained in Hoogendoorn et al.,[34] a Weibull distribution was used to simulate time to death (at baseline) for

Table 1 Regression Equations used in the COPD Model[a]

Outcome and Predictors for Each Regression Equation in the COPD Model	R Function Used to Fit the Regression Equation	Use in the Model
Time to exacerbation ~ stable baseline characteristics + age + lung function + physical activity + previous exacerbations + severity previous exacerbations + disease-specific quality of life Time to pneumonia ~ stable baseline characteristics + age + FEV1 percentage predicted + severity previous exacerbations + physical activity + exercise capacity + symptoms + disease-specific quality of life	Regression for parametric survival models (*survreg* with Weibull distribution).[40] The 2 parameters of a Weibull curve (shape and scale) for each patient are estimated by filling in the patient and disease characteristics of that specific patient in the corresponding equation. Based on these (shape and scale) parameters, an individual survival curve for a patient is constructed.	Estimate time to event. A random value is drawn from the corresponding individual Weibull curve to determine the time to event (exacerbation, pneumonia or death) for a patient.
Probability that exacerbation is severe ~ stable baseline characteristics + age + lung function + physical activity + previous exacerbations + severity of previous exacerbations + disease-specific quality of life Probability of hospitalization due to pneumonia ~ stable baseline characteristics + age Probability of shortness of breath ~ stable baseline characteristics + time + age + lung function + severity previous exacerbations + physical activity + exercise capacity + previous shortness of breath Probability of cough/sputum ~ stable baseline characteristics + time + age + lung function + severity of previous exacerbations + physical activity + exercise capacity + previous cough/sputum	Generalized linear mixed-effects model with binomial family (*glmer* with binomial family).[41] The log odds of a binary outcome for each patient are estimated by filling in the patient and disease characteristics of that specific patient in the corresponding equation. In a second step, the log odds are transformed into a probability using the following formula: exp(log odds) / (1 + exp(log odds)).	Predict binary outcomes. A random value is drawn from a Bernoulli distribution with the corresponding individual probability to have a binary outcome (severe exacerbation, pneumonia leading to hospitalization, patient is presented with shortness of breath, patient is presented with cough/sputum) for a patient. If the random drawn value is 1, the outcome is assumed to be present for a patient.
Lung function ~ stable baseline characteristics * time + age at baseline + lung function at baseline + previous exacerbations + severity of previous exacerbations Exercise capacity ~ stable baseline characteristics + age + lung function + physical activity + exercise capacity + previous exacerbations Physical activity ~ stable baseline characteristics + time + age + lung function + physical activity + exercise capacity + symptoms + disease-specific quality of life Disease-specific quality of life ~ stable baseline characteristics + time + age + previous disease-specific quality of life + lung function + severity previous exacerbations + physical activity + exercise capacity + symptoms + pneumonia	Linear mixed-effects model with random intercept (*lme*).[42] Continuous outcomes over time are estimated for each patient by filling in the patient and disease characteristics of that specific patient in the corresponding equation.	Estimate continuous lung function over time (defined as FEV1 in liters). Estimate continuous exercise capacity (defined as treadmill test in seconds). Estimate continuous physical activity (defined as St. George's Respiratory Questionnaire [SGRQ] activity score in points between 0 and 100). Estimate continuous disease-specific quality of life (defined as SGRQ total score in points between 0 and 100).
Health care use ~ stable baseline characteristics + age + lung function + severity previous exacerbations + physical activity + exercise capacity + symptoms + disease-specific quality of life	Negative binomial generalized linear model (*glm.nb*)[43]	Estimate health care use (the number of general practitioner and specialist visits per year).

COPD, chronic obstructive pulmonary disease; FEV1, forced expiratory volume in 1 second.

[a]Stable baseline characteristics = sex, body mass index, smoking status, number of pack-years smoked, heart failure, other cardiovascular disease, reversibility, diabetes, depression, asthma, emphysema, inhaled corticosteroids (ICS) use, and high eosinophils. For further details including definitions of the patient characteristics and results from estimating the equations (e.g., regression coefficients), we refer to Hoogendoorn et al.[34]
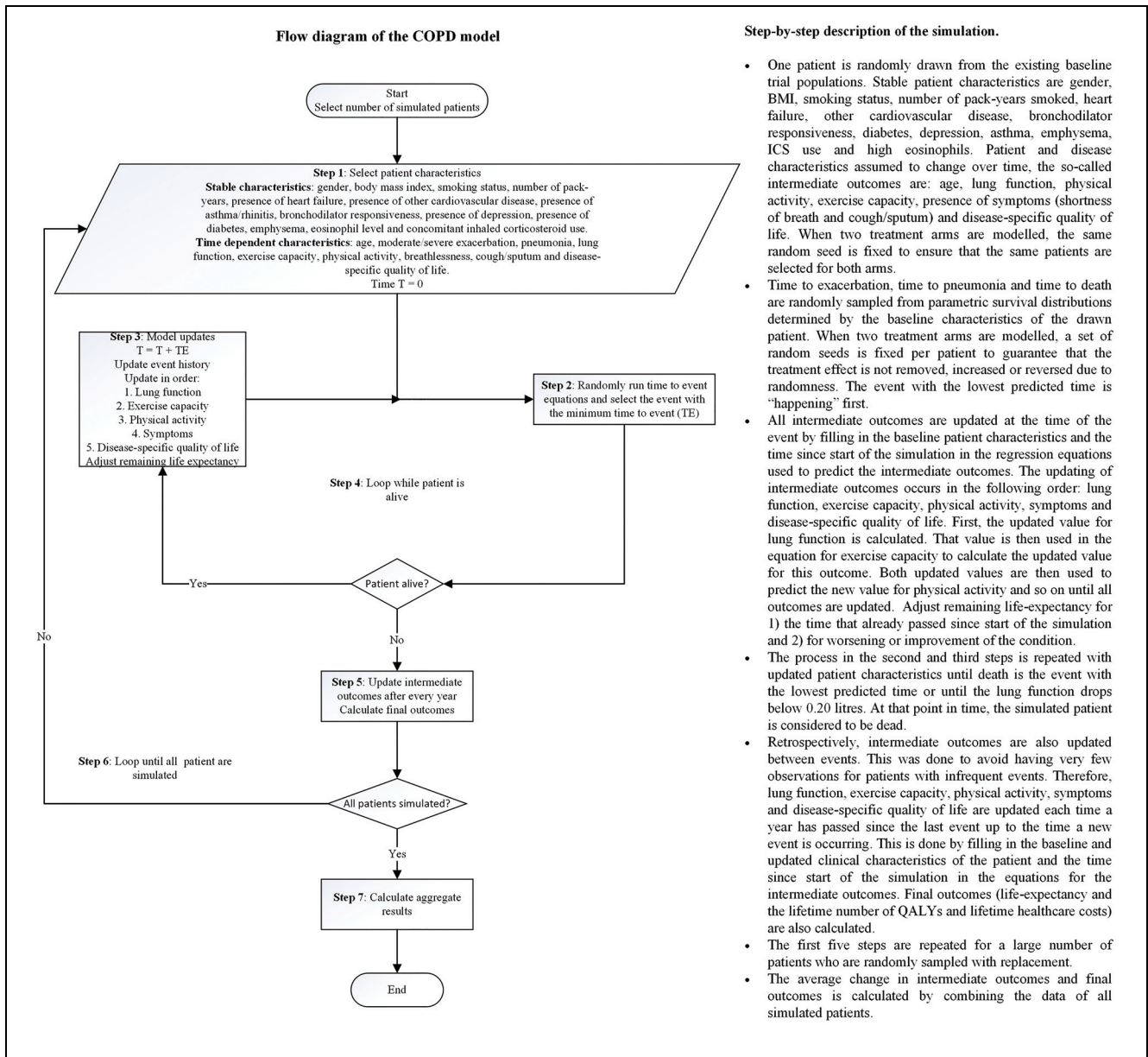
Figure 1 Flow diagram and step-by-step description of the chronic obstructive pulmonary disease (COPD) model. QALY, quality-adjusted life year.

patients in the COPD model. The parameters (shape and scale) of the Weibull distribution function were estimated using the equation shown in Table 1, which resulted in the patient's life expectancy predicted at baseline. During the simulation, life expectancy is affected by the occurrence of events and the changes in age and intermediate outcomes. To predict remaining life expectancy at the time of an event, updated characteristics are filled in the equation. If, at the time of an event, a new remaining life

expectancy was randomly drawn, this could result in inconsistent results (e.g., a patient whose condition worsened after having a severe exacerbation could be randomly assigned a life expectancy higher than that before having the exacerbation). To avoid this issue, the random seed 2 in the clinical history function shown in Figure 2 was used. However, this was not sufficient since inconsistencies between a worsening condition and increased life expectancy (and vice versa) were observed, which called

Pseudo code for the clinical history function.

```
COPD_model_simulation_clinical_history <- function(baseline characteristics, tre
atment effect parameters, patient size, PSA (yes/no), random seed){

        Read regression coefficients
        Read data with patient characteristics
        Set random seed #1
        IF (PSA = yes) THEN{
            Sample regression coefficients
            Sample treatment effect parameters
            }
        Select the sample (with replacement) of patients
        Create patient history table (now empty)

        Set patient index = 1
        FOR (patient index = 1 TO patient size){
            Select current patient
            Predict continuous patient characteristics at baseline
            Set random seed #2
            Draw life expectancy at baseline
            Set event index = 1
            while(patient = alive){
                Set random seeds #3 (one for each event while patient alive)
                Sample time to exacerbation
                Sample time to pneumonia
                Select event with minimum time
                Update patient characteristics
                Adjust remaining life expectancy
                Event index + 1
                } End while loop and move to another patient
            Save patient history
            Patient index + 1
            } End for loop in number of patients

        Update intermediate outcomes after every year
        Calculate aggregated results
        Return full patients histories
} End clinical history function
```

Pseudo code for the PSA function.

```
COPD_model_PSA <- function(outer loop, inner loop, treatment effect parameters,
unit costs, resource use, discount rates, retirement age){
        Create (empty) matrix to store PSA results
        Set PSA index = 1
        FOR (PSA index = 1 to outer loop){
            Simulate clinical history(random seed = PSA index)
            Calculate costs
            Calculate QALYs
            Calculate average results per PSA iteration
            }
        Calculate average results across all PSA iterations
        Return PSA results
}
```

Figure 2 Pseudo-code of the chronic obstructive pulmonary disease (COPD) model.

for an adjustment of the life expectancy. When life expectancy is updated at the time of an event, the outcome of the time to death equation reflects the remaining life expectancy at baseline should the patient have had the updated characteristics at baseline (because regression coefficients were estimated based on baseline data). Therefore, the remaining life expectancy needs to be corrected for 1) the time that already had passed since the start of the simulation and 2) for worsening or improvement of the condition. Two simulated patients are used to illustrate how this was done. The first patient's baseline characteristics are shown in the first row of Table 2. The remaining life expectancy at baseline (RLE_t0) was predicted by filling in these characteristics

Table 2 Example of Two Simulated Patient Histories

Patient ID	Time, y	Age, y	FEV$_1$	Severe exacerbation (Yes = 1)	Moderate exacerbation (Yes = 1)	Exercise capacity, s	SGRQ Activity Score	SGRQ Total Score	Cough/Sputum (Yes = 1)	Breathlessness (Yes = 1)	Dead
1	0.00	73.00	1.22	0	1	376.58	57.84	40.35	0	1	0
	1.33	74.33	1.18	0	1	410.95	57.60	36.17	0	0	0
	2.20	75.20	1.15	0	1	436.07	56.13	38.19	1	0	0
	6.39	79.39	1.02	0	1	448.21	59.30	45.60	1	1	0
	11.87	84.87	0.85	0	1	445.95	67.75	48.39	1	0	0
	11.87	84.87	0.86	0	0	480.74	69.81	48.03	1	0	1
2[a]	0.00	61.00	1.08	0	0	366.30	57.77	48.25	1	1	0
	1.00	62.00	1.04	0	0	374.38	59.93	49.20	1	1	0
	2.00	63.00	0.99	0	0	377.66	61.76	50.56	1	1	0
	3.00	64.00	0.95	0	0	377.24	63.61	51.97	1	1	0
	3.05	64.05	0.93	0	1	331.39	58.21	46.61	1	0	0
	. . .										
	13.74	74.74	0.47	0	0	165.99	84.17	67.69	1	1	0
	14.68	75.68	0.42	0	0	163.02	81.14	62.28	0	1	1

FEV$_1$, forced expiratory volume in 1 second; SGRQ, St. George's Respiratory Questionnaire.
[a]The complete clinical history simulated for this patient is not shown in this table.

in the time-to-death equation. In the example, the *expected* life expectancy at baseline was 11.43 years. In the beginning of the simulation, a *random* time to death at baseline was drawn (12.36 years). This value was used as reference for the remaining simulation. At the time of the first event (1.33 years), all parameters were updated (second row in Table 2), and the updated remaining life expectancy was calculated using the following formula:

Remaining life expectancy at first event (RLE_t1) = (RLE_t0 – time passed until the first update (t1)) * (Expected life expectancy at baseline using updated values at t1 / Expected life expectancy at baseline using baseline values).

The expected life expectancy at baseline had that patient had the characteristics in the second row of Table 2 was 10.82 years, which was lower than the value obtained with the baseline characteristics (11.43 years), reflecting the patient's worsened condition. The remaining life expectancy after the first event was (12.36 – 1.33) * (10.82/11.43) = 10.44 years. Thus, the remaining life expectancy after the first event (10.44 years) plus the time that had passed until the first event happened (1.33 years) is 11.77 years, which is lower than the remaining life expectancy estimated at baseline (12.36 years). This difference (0.59 years) is the result of the patient's worsened condition. For the second patient in Table 2, the expected life expectancy at baseline was 10.63 years, and the random time to death at baseline drawn in the

beginning of the simulation was 15.21 years. At the time of the first update (at 1 year, so no event occurred), all parameters were updated (eighth row in Table 2) and the updated remaining life expectancy was 10.81 years, which was higher than the value obtained with the baseline characteristics, reflecting the patient's improved condition. The remaining life expectancy after year 1 was (15.21 – 1) * (10.81/10.63) = 14.45 years. Thus, the remaining life expectancy after the first year was 15.45 years, which is higher than the remaining life expectancy estimated at baseline (15.21 years). The difference of 0.24 years is the result of the patient's improved condition. This adjustment was repeated for all patients until the patient's death. The ratio in the above formula reflects a factor to correct for improvement or deterioration in the patient's health status over time.

Challenge 3: Remove stochastic uncertainty from treatment effectiveness

Why is this challenge important? Stochastic uncertainty refers to the random variability in the model outcomes between identical patients. Even though, by fixing a random seed, it is possible to ensure that patients are the same for different simulations, the results of these still include stochastic uncertainty. Due to the (random) sampling of times to event, every time a simulation is run, results could be different even if the same patients are selected, simply because different event times can be drawn. Because of this, the modeled treatment effects for an intervention can be masked, which should not happen.

What solution was implemented? Suppose that we want to model a new COPD intervention whose main effect consists of delaying the occurrence of exacerbations. The time to first exacerbation is determined by the same Weibull curve in both arms because at the beginning of the simulation, the same patient (with the same characteristics) starts in both arms. In the base-case arm, the time to first exacerbation is randomly sampled from the corresponding Weibull curve. In the new intervention arm, the time to first exacerbation is simply the time sampled in the base-case increased by, for example, 15%. Because in the intervention arm, the first exacerbation was postponed, the patient's age and intermediate outcomes after the first exacerbation (i.e., when the time to next event is calculated) are not the same as in the base-case arm. Consequently, the time to second exacerbation is not determined by the same Weibull curves. Appendix 3 (available online) provides a more detailed explanation. If 2 Weibull curves were used to sample the time to next exacerbation independently, the sampled times could be inconsistent with the delay in exacerbation due to stochastic uncertainty. Drawing a low random value for the time to next exacerbation in the intervention arm and a high random value in the comparator arm would result in an intervention being less effective than a comparator. To ensure consistency, the set of random seeds 3 in the clinical history function shown in Figure 2 was fixed per patient. These seeds guarantee that the treatment effect is not removed, increased, or reversed due to randomness. This approach always results in a positive effect of the intervention unless, by extending the time to exacerbation, a pneumonia or death occurs.

Challenge 4: Remove heterogeneity and stochastic uncertainty from probabilistic sensitivity analysis

Why is this challenge important? Probabilistic sensitivity analysis (PSA) assesses the magnitude of parameter uncertainty in HE models[44] and, in patient-level models, is implemented as a double loop: the number of iterations in the PSA (outer loop) and the number of patients per PSA iteration (inner loop).[7,45,46] Loop sizes should be determined in such a way that the PSA results are stable. Stochastic uncertainty and patient heterogeneity should not be included in a PSA.[46,47] Input parameters should be the same across treatment arms. That way, the difference between the 2 arms in the PSA only results from the application of a treatment effect.

What solution was implemented? Parameter uncertainty in the COPD model is the uncertainty around the estimation of the regression coefficients and the treatment effect parameters included in the model. The estimated regression coefficients are provided in Appendix 4 (available online). Uncertainty around treatment effect parameters (e.g., percentage increase in the time to exacerbation) is based on the available efficacy or effectiveness data for new interventions. Appendix 5 (available online) provides details about the estimation of the PSA loop sizes in the COPD model. The PSA function in the COPD model calls other model functions multiple times and calculates average results per iteration. The pseudo-code for the PSA function is shown in Figure 2. Every time the clinical history function is called, the random seed required as input parameter is changed with the PSA index. That way, the parameters drawn in the PSA are different per iteration but the same across treatment arms, and the difference in results is only due to the application of a treatment effect. The uncertainty around the regression coefficients was addressed by assuming a multivariate normal distribution with parameters equal to the estimated regression coefficients and the estimated covariance matrices shown in Appendix 4 (available online). In each PSA iteration, random draws are taken from these multivariate distributions to get a new set of coefficients for each equation, which are then combined with the characteristics of the patients included in the PSA. The uncertainty around the treatment effect parameters was included by assuming a certain deviation (%) from the assumed mean value and then sampling from a uniform distribution. No uncertainty was included around treatment costs.

Cost-Effectiveness Scenario Analyses

To illustrate the potential impact of each type of uncertainty on model outcomes, we ran the COPD model for several hypothetical scenarios. In the base-case (deterministic) scenario, all solutions described above were implemented. This scenario was run for 1000 individual patients treated with a common medication for COPD patients (tiotropium), which acts as the comparator arm in the model. For the intervention arm, we used a hypothetical treatment that delays the time to exacerbation by 15% and is 50% more expensive compared to the base-case. Note that this treatment effect is not unrealistic. In the base-case scenario, an increase in time to exacerbation with 15% resulted in a reduction in the exacerbation rate with a rate ratio of 0.86, which is within the range of rate ratios observed for several treatments in large COPD trials.[35–39] A PSA was also conducted as explained in Challenge 4 and by assuming an additional ±10% variation to the delayed time to exacerbation. The size of the outer and inner loops was 300 and 100, respectively. According to the calculations

shown in Appendix 5 (available online), the PSA loop sizes are large enough to provide stable results. Four additional scenarios (described below) were also run to illustrate what might occur when the solutions described above are not implemented.

Scenario 1: Differences in patient heterogeneity between the intervention and control arms not removed. Scenario 1 was run using different random seeds for selecting patients in the intervention and the comparator arm. Therefore, the 1000 patients in the intervention arm were not the same as the 1000 patients in the comparator arm. The random seeds used to sample the times where the events occurred were the same across treatment arms. Hence, differences in results between the base-case scenario and scenario 1 can be attributed to differences in patient heterogeneity between intervention and control arms.

Scenario 2: Unadjusted remaining life expectancy after the occurrence of an event. In this scenario, we considered for both arms the same 1000 patients and the same random seeds used to sample time to event as in the base-case scenario. However, in scenario 2, the remaining life expectancy after events occurred in the simulation was not adjusted, as described in Challenge 2. The differences in results between scenario 2 and the base-case scenario can be thus attributed to the alternative modeling assumptions regarding life expectancy.

Scenario 3: Stochastic uncertainty not removed from treatment effectiveness. In scenario 3, the same 1000 patients were selected for the intervention and the comparator arm. However, random seeds were not fixed when sampling time to event in the intervention arm. Thus, differences in results between the base-case scenario and scenario 3 can be attributed to the stochastic uncertainty associated to the random drawing of time to event.

Scenario 4: Heterogeneity and stochastic uncertainty included in probabilistic sensitivity analysis. As explained in Challenge 4, in the PSA, an additional set of random seeds is required to ensure that the parameters drawn in the PSA are different per iteration but the same across treatment arms. In scenario 4, we run a new PSA where in each iteration, the same patients and the same random seeds used to sample time to event as in the base-case scenario were selected. However, the PSA here was run without fixing the PSA-specific random seeds, and

therefore, the results include stochastic uncertainty and patient heterogeneity, which is methodologically incorrect.[45,46] Thus, only the PSA in the base-case scenario adequately reflects parameter uncertainty.

Results

In the base-case scenario, the 4 forms of uncertainty were implemented in the way it was considered to be correct. The base-case results are thus believed to be appropriate for decision making on a hypothetical intervention, which was assumed to delay time to exacerbation by 15% and to be 50% more expensive. This intervention generated 0.0728 incremental QALYs with higher incremental costs of €1449, resulting in an incremental cost-effectiveness ratio (ICER) of €19,904 per QALY gained. At a common threshold ICER of €20,000 per QALY gained, the new intervention could be deemed as cost-effective, even though the ICER is close to the threshold.

In scenario 1, as shown in Table 3, the ICER was approximately €10,000 larger than the base-case ICER, falling thus above the threshold ICER of €20,000 per QALY gained. Therefore, by selecting different patients per treatment arm, we moved from a situation of a "borderline" ICER in the base-case to a situation where the intervention is not cost-effective in scenario 1.

In scenario 2, remaining life expectancy after events was not adjusted, as explained in Challenge 2. As shown in Table 4, both the total costs and total QALYs were similarly increased across treatment arms compared to the base-case. Because of this, the ICER in scenario 2 was €19,943, thus in line with the base-case ICER and still below the €20,000 threshold. This increase in total costs and QALYs was mostly caused by an extended life expectancy of approximately 1.4 years. In scenario 2, patients live longer because the model is not corrected for a worsening in the COPD condition.

In scenario 3, the same patients were selected in the intervention and the comparator arm, but no random seeds were fixed to sample time to events in the intervention arm. As a result, the new intervention had similar incremental costs, but the incremental QALYs were very low. Consequently, the ICER was almost 14 times higher compared to the base-case (see Table 3). Thus, by randomly sampling time to events in the intervention arm that were shorter by chance, we moved from a "borderline" ICER in the base-case to a very high ICER in scenario 3, in which the new intervention would not be deemed as cost-effective.

PSA results are also shown in Table 3. The ICER obtained from the PSA base-case analysis was €22,258,

Table 3 Example of COPD Model Results Affected by Patient Heterogeneity and Stochastic Uncertainty and PSA With and Without Fixing PSA-Specific Random Seeds[a]

Scenario	Technologies	Total Costs (€)	Total QALYs	Incremental Costs (€)	Incremental QALYs	ICER (€)
Base-case (deterministic)	Comparator	€17,567	5.7717			
	Intervention	€19,016	5.8445	€1449	0.0728	€19,904
Scenario 1	Intervention	€19,184	5.8260	€1617	0.0543	€29,779
Scenario 3	Intervention	€19,004	5.7770	€1437	0.0053	€271,132
Base-case (PSA)	Comparator	€17,147	5.7163			
	Intervention	€18,580	5.7806	€1432	0.0644	€22,258
Scenario 4	Comparator	€17,462	5.7468			
	Intervention	€18,824	5.7812	€1363	0.0343	€39,677

ICER, incremental cost-effectiveness ratio; PSA, probabilistic sensitivity analysis; QALYs, quality-adjusted life years.
[a]Scenario 1: Different random seed for selecting patients for the intervention and the comparator (heterogeneity). Scenario 3: Same patients as in base-case, but no random seeds were fixed for sampling time to events (stochastic uncertainty). Scenario 4: PSA with no random seeds fixed.

Table 4 Results for a Simulated Cohort of Patients With and Without Adjusting Remaining Life Expectancy (Structural Uncertainty)[a]

Scenario	Total Costs (€)	Total QALYs	Mean Life Expectancy, y	Mean Exacerbation Rate per Year
Base-case (comparator): adjusting RLE	€17,567	5.7717	11.45	0.674
Scenario 2 (comparator): without adjusting RLE	€20,281	6.0499	12.84	0.712
Difference (comparator)	€2714	0.2782	1.39	0.038
Base-case (intervention): adjusting RLE	€19,016	5.8445	11.64	0.585
Scenario 2 (intervention): without adjusting RLE	€21,675	6.1198	13.06	0.614
Difference (intervention)	€2659	0.2753	1.42	0.029

ICER, incremental cost-effectiveness ratio; QALYs, quality-adjusted life years; RLE, remaining life expectancy.
[a]Base-case ICER: €19,904; scenario 2 ICER: €19,943.

whereas the probabilistic ICER in scenario 4 was €39,677. Even though the probabilistic base-case ICER was above the €20,000 threshold, it might be deemed as a "borderline" ICER. However, in scenario 4, the probabilistic ICER is clearly above that threshold. Furthermore, when the PSA outcomes were plotted in the cost-effectiveness (CE) plane, it was clear that in scenario 4, the uncertainty was much larger (Figure 3). In scenario 4, the PSA was run without fixing PSA-specific random seeds, which caused results to be scattered over all quadrants of the CE plane. The cost-effectiveness acceptability curve (CEAC) in the base-case showed the common increasing shape when the PSA outcomes are mostly in the northeastern quadrant of the CE plane, while in scenario 4, the CEAC flattened quickly. However, in the PSA base-case, the hypothetical intervention had approximately a 40% probability of being cost-effective at a threshold of €20,000 per QALY, whereas in scenario 4, this was 46%. Looking at these probabilities in isolation

can be misleading since based on Figure 3, it seems clear that uncertainty is a great concern in scenario 4.

Discussion

In this article, we have presented 4 challenges associated with modeling uncertainty that were encountered during the implementation of a DES COPD model but that can be applicable to DES models in general. The first challenge was to remove the differences in patient heterogeneity between the intervention and control groups. The solution proposed in the COPD model consisted of fixing a random seed (see Figure 2, seed 1). From this challenge, we learned that in patient-level models, patient heterogeneity can lead to an erroneous interpretation of treatment effects. This was illustrated in scenario 1, where the new (hypothetical) intervention increased the ICER by approximately 50%. However, this was not caused by the assumed treatment effect, as it should be,
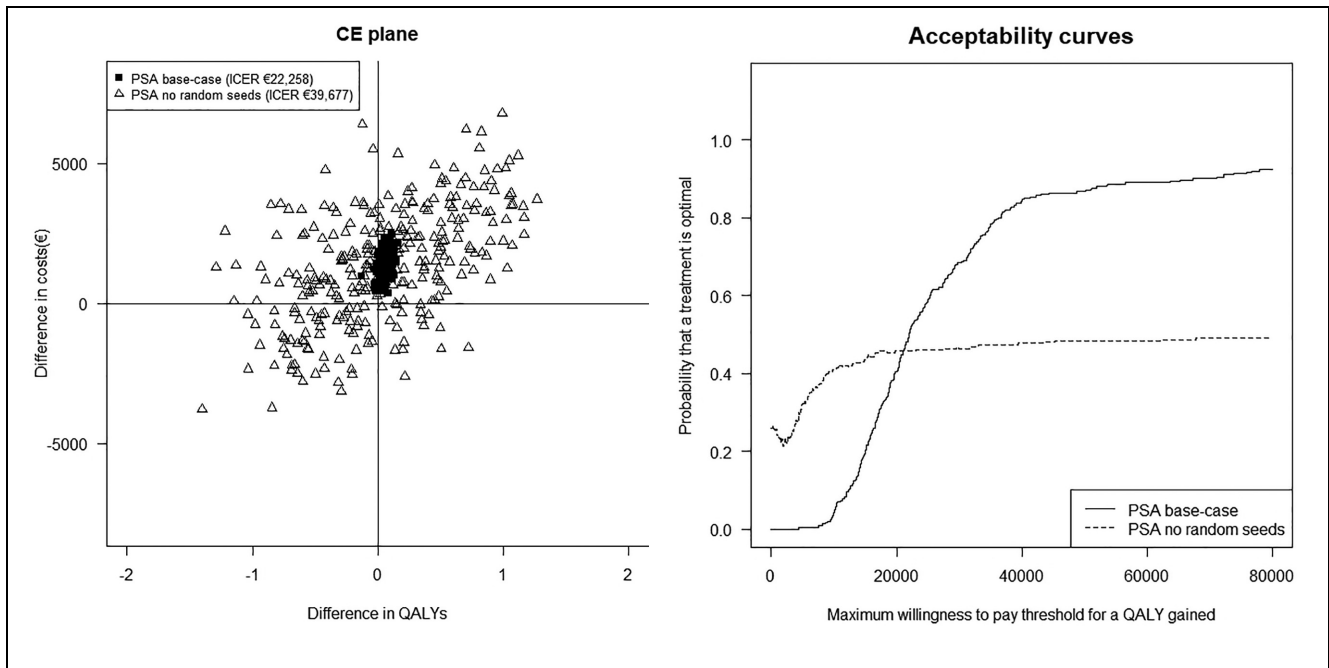
Figure 3 Example of probabilistic sensitivity analysis (PSA) results with and without fixing PSA-specific random seeds. ICER, incremental cost-effectiveness ratio; QALYs, quality-adjusted life years.

but because different patients were selected for each treatment arm. The new intervention was less effective for the patients selected in scenario 1 than for those selected in the base-case scenario. Thus, the differences in the model results were attributed to differences in heterogeneity rather than to differences in effectiveness between treatments. The second challenge consisted of adjusting the remaining life expectancy after the occurrence of a COPD-related event (i.e., exacerbation or pneumonia). The solution proposed in the COPD model was to fix another random seed (see Figure 2, seed 2) and to correct the remaining life expectancy for 1) the time that already had passed since start of the simulation and 2) for worsening or improvement of the condition, as explained in Challenge 2. From this challenge, we learned that it is important to assess the impact not only on HE outcomes but also on clinical outcomes. Face validity of the clinical outcomes can be one of the reasons for preferring one modeling assumption over a plausible set of alternatives. By making these changes, the model results became more valid, but also an additional element of structural uncertainty was introduced. The third challenge was to remove stochastic uncertainty from treatment effectiveness. The solution proposed in the COPD model was to fix a set of random seeds per patient (see Figure 2, seed 3). From this challenge, we learned that in

stochastic models, random chance can also lead to an erroneous interpretation of treatment effects. As an example, scenario 3 resulted in an ICER that was almost 14 times higher than the base-case ICER. Overall, this was due to shorter times (to event) sampled in the intervention arm. However, this was not caused by the assumed treatment effect, as it should be, but because the times (to event) randomly sampled for the intervention arm in scenario 3 were by chance shorter than those sampled in the intervention arm in the base-case scenario. The fourth and final challenge addressed in this article was to remove heterogeneity and stochastic uncertainty from the PSA. In the COPD model, the PSA function calls the other model functions multiple times and calculates average results per iteration. Thus, the solution proposed in the COPD model was to use as input parameter for the clinical history function a different random seed (which changes with the PSA index) every time that function was called (see Figure 2). From this challenge, we learned that in probabilistic (as opposed to deterministic) models, the uncertainty associated with the model results can be misrepresented when the input parameters drawn in the PSA are not the same across treatment arms. This was illustrated in scenario 4, where the uncertainty around the model results was larger than in the base-case. In both scenarios, the model had the same number of

input parameters, the same probability distributions, the same patients, and the same random seeds used to draw time to event. However, for each individual patient run in the PSA, the input parameters were different for each treatment arm: they were randomly drawn regardless of the treatment arm, which resulted in model outcomes scattered all over the CE plane.

Random sampling was presented as a key concept in DES models, and its importance as a main differentiator from common cohort models was highlighted. In the COPD model used as an example in this article, random sampling was done to select the patient population, to calculate time to events and as a part of the PSA. We explained that it was necessary to control the random number generation process to ensure the consistency of the results. It is important to emphasize that in TSD15, the use of random seeds is explained in the context of replicability of results.[7] In the COPD model, random seeds were also needed for consistency of the results (even if the model was run just once), as explained throughout the Methods section of this article. All these seeds, after being fixed, guaranteed that the treatment effect was not removed/increased or reversed due to patient heterogeneity and stochastic or parameter uncertainty, which can have a great impact on the model results, as shown in the Results section. This, however, does not imply that using fixed random seeds in the way previously described always results in positive treatment effects. For example, based on the uncertainty interval around the treatment effect considered in the PSA, negative effects can also occur.

We must acknowledge the presence of structural uncertainty in all decision models and that this type of uncertainty may have a considerable impact on the model results. While some assumptions can be tested in a systematic way (like the choice of parametric survival distributions), testing other assumptions is not so straightforward or simply not possible/feasible. An example of this could be the reestimation of intermediate (nonevent) outcomes in the COPD model. Time to event was calculated at baseline and each time an event occurred, but it was not reestimated after each year. The main reason for doing the retrospective update was to report intermediate outcomes on annual basis as this is the most common way to report them (e.g., annual decline in lung function). The main implication of this approach was for the calculation of QALYs. In the COPD model, utilities are calculated as a function of the SGRQ total score. By having SGRQ updated every year, annual QALYs can also be calculated, which, in turn, results in a more accurate estimation of the QALYs accrued over the simulated lifetime. We assumed that the time to next event was

determined by the patient characteristics at baseline (for the first event) and the patient characteristics at the time of an event for the next events. We believe this is the common way to perform DES, where time to events are simulated. Furthermore, the most important predictor for an exacerbation is a previous exacerbation, as can be seen from the estimated regression coefficients presented in Appendix 4, Table A4.1 (available online). If we had updated events every year, we would have adopted a different approach, for example, to calculate the annual probability of having an exacerbation. This would not be a time-to-event model and would result in different outcomes (life expectancy, QALYs). This alternative way of modeling, if appropriately implemented, would also be valid and illustrates yet another example of structural uncertainty.

Conclusions

Modeling uncertainty is crucial in all health economic decision models but even more so in DES models, where all the possible types of uncertainty (i.e., patient heterogeneity, stochastic, parameter and structural uncertainty) apply. Because DES models are usually complex and less often used, it is especially important to be as detailed as possible in the description of the methodology used to model uncertainty. In our experience, this is not always the case. With this in mind, we believe this article can be a valuable addition to the literature on DES. We have reported the methods used in the model with a great extent of detail and as clearly as we could (the latter is of course subjective and depends very much on the reader). By doing this, we can help other model developers, who are likely to face similar challenges to those described here. We have provided examples, using our COPD model as case study, to illustrate the issues described in the Methods section of the article. These examples can help the reader to get a better understanding of the issues presented there. We have included a link to the full R code of the model where readers can see how the proposed solutions were implemented. For readers who are not familiar with R, we have provided the model's pseudo-code, which can be used as a standalone tool to understand how the flow of the code works without knowing the specifics of the R language. To the best of our knowledge, this level of detail is not common in the HE modeling literature. Finally, we also hope to encourage other model developers to do the same, which in the long term will help increase the transparency of future DES models. We think this is the way forward until the HE community is ready to accept more transparent approaches, like using open-source models.

ORCID iD

Isaac Corro Ramos (iD) https://orcid.org/0000-0002-1294-8187

Supplemental Material

References

1. Redekop WK, Mladsi D. The faces of personalized medicine: a framework for understanding its meaning and scope. *Value Health*. 2013;16:S4–9.
2. Zafari Z, Bryan S, Sin DD, Conte T, Khakban R, Sadatsafavi M. A systematic review of health economics simulation models of chronic obstructive pulmonary disease. *Value Health*. 2017;20(1):152–62.
3. Di Tanna GL, Bychenkova A, O'Neill F, et al. Evaluating cost-effectiveness models for pharmacologic interventions in adults with heart failure: a systematic literature review. *Pharmacoeconomics*. 2019;37(3):359–89.
4. Henriksson M, Jindal R, Sternhufvud C, Bergenheim K, Sörstadius E, Willis M. A systematic review of cost-effectiveness models in type 1 diabetes mellitus. *Pharmacoeconomics*. 2016;34(6):569–85.
5. Schiller-Frühwirth IC, Jahn B, Arvandi M, Siebert U. Cost-effectiveness models in breast cancer screening in the general population: a systematic review. *Appl Health Econ Health Policy*. 2017;15(3):333–51.
6. Miller JD, Foley KA, Russell MW. Current challenges in health economic modeling of cancer therapies: a research inquiry. *Am Health Drug Benefits*. 2014;7(3):153–62.
7. Davis S, Stevenson M, Tappenden P, Wailoo A. NICE DSU Technical Support Document 15: cost-effectiveness modelling using patient-level simulation. 2014. Available from: http://nicedsu.org.uk/technical-support-documents/technical-support-documents
8. Karnon J, Stahl J, Brennan A, et al. Modeling using discrete event simulation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-4. *Value Health*. 2012;15:821–7.
9. Caro JJ, Möller J, Getsios D. Discrete event simulation: the preferred technique for health economic evaluations? *Value Health*. 2010;13(8):1056–60.
10. Brennan A, Chick SE, Davies R. A taxonomy of model structures for economic evaluation of health technologies. *Health Econ*. 2006;15:1295–310.
11. Stahl JE. Modelling methods for pharmacoeconomics and health technology assessment: an overview and guide. *Pharmacoeconomics*. 2008;26:131–48.
12. Karnon J, Brown J. Selecting a decision model for economic evaluation: a case study and review. *Health Care Manag Sci*. 1998;1:133–40.
13. Caro JJ, Briggs AH, Siebert U, Kuntz KM. Modeling good research practices—overview: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force–1. *Med Decis Making*. 2012;32(5):667–77.
14. Caro JJ, Migliaccio-Walle K, O'Brien JA, et al. Economic implications of extended-release metoprolol succinate for heart failure in the MERIT-HF Trial: a US perspective of the MERIT-HF Trial. *J Cardiac Failure*. 2005;11(9):647–56.
15. Degeling K, Franken MD, May AM, et al. Matching the model with the evidence: comparing discrete event simulation and state-transition modeling for time-to-event predictions in a cost-effectiveness analysis of treatment in metastatic colorectal cancer patients. *Cancer Epidemiol*. 2018;57:60–7.
16. Geitona M, Stamuli E, Giannakodimos S, et al. Lacosamide as first-line treatment option in focal epilepsy: a cost-utility analysis for the Greek healthcare system. *J Med Econ*. 2019;22(4):359–64.
17. Ibarrondo O, Álvarez-López I, Freundlich F, et al. Probabilistic cost-utility analysis and expected value of perfect information for the Oncotype multigenic test: a discrete event simulation model. *Gac Sanit*. 2020;34(1):61–8.
18. Jiao F, Wan EYF, Fung CSC, et al. Cost-effectiveness of a primary care multidisciplinary Risk Assessment and Management Program for patients with diabetes mellitus (RAMP-DM) over lifetime. *Endocrine*. 2019;63(2):259–69.
19. Larrañaga I, Stafylas P, Fullaondo A, Apuzzo GM, Mar J. Economic evaluation of an integrated health and social care program for heart failure through 2 different modeling techniques. *Health Serv Res Manag Epidemiol*. 2018;5:2333392818795795.
20. Le QA, Hay JW, Becker R, Wang Y. Cost-effectiveness analysis of sequential treatment of abaloparatide followed by alendronate versus teriparatide followed by alendronate in postmenopausal women with osteoporosis in the United States. *Ann Pharmacother*. 2019;53(2):134–43.
21. Lee D, Wilson K, Akehurst R, et al. Cost-effectiveness of eplerenone in patients with systolic heart failure and mild symptoms. *Heart*. 2014;100:1681–7.
22. McEwan P, Poole CD, Tetlow T, Holmes P, Craig J. Currie Evaluation of the cost-effectiveness of insulin glargine versus NPH insulin for the treatment of type 1 diabetes in the UK. *Curr Med Res Opin*. 2007;23(suppl 1):S7–S19.
23. Mueller E, Maxion-Bergemann S, Gultyaev D, et al. Development and validation of the economic assessment of glycemic control and long-term effects of diabetes (EAGLE) model. *Diabetes Technol Ther*. 2006;8(2):219–36.
24. Noon KM, Montgomery SM, Adlard NE, Kroes MA. When does economic model type become a decisive factor in health technology appraisals? Learning from the expanding treatment options for relapsing-remitting multiple sclerosis. *J Med Econ*. 2018;21(10):983–92.
25. Tarhini A, Benedict A, McDermott D, et al. Sequential treatment approaches in the management of BRAF wild-type advanced melanoma: a cost-effectiveness analysis. *Immunotherapy*. 2018;10(14):1241–52.
26. Tarhini A, McDermott D, Ambavane A, et al. Clinical and economic outcomes associated with treatment sequences in

patients with BRAF-mutant advanced melanoma. *Immunotherapy*. 2019;11(4):283–95.

27. Valentine WJ, Pollock RF, Saunders R, Bae J, Norrbacka K, Boye K. The prime diabetes model: novel methods for estimating long-term clinical and cost outcomes in type 1 diabetes mellitus. *Value Health*. 2017;20(7):985–91.

28. Wang HI, Roman E, Crouch S, et al. A generic model for follicular lymphoma: predicting cost, life-expectancy, and quality-adjusted-life-year using UK population-based observational data. *Value Health*. 2018;21(10):1176–85.

29. Burr J, Botello Pinzon P, Takwoingi Y, Hernandez R, Vazquez-Montes M. Surveillance for ocular hypertension: an evidence synthesis and economic evaluation. *Health Technol Assess*. 2012;16(29):1–271.

30. Grant SW, Sperrin M, Carlson E, et al. Calculating when elective abdominal aortic aneurysm repair improves survival for individual patients: development of the Aneurysm Repair Decision Aid and economic evaluation. *Health Technol Assess*. 2015;19(32):1–154.

31. Simpson EL, Stevenson M D, Rawdin A, Papaioannou D. Thrombophilia testing in people with venous thromboembolism: systematic review and cost-effectiveness analysis. *Health Technol Assess*. 2009;13(2):iii, ix–x, 1–91.

32. McManus E, Turner D, Gray E, Khawar H, Okoli T, Sach T. Barriers and facilitators to model replication within health economics. *Value Health*. 2019;22(9):1018–25.

33. Degeling K, Koffijberg H, IJzerman MJ. A systematic review and checklist presenting the main challenges for health economic modeling in personalized medicine: towards implementing patient-level models. *Exp Rev Pharmacoecon Outcomes Res*. 2017;17(1):17–25.

34. Hoogendoorn M, Corro Ramos I, Baldwin M, Gonzalez-Rojas Guix N, Rutten-van Mölken MPMH. Broadening the perspective of cost-effectiveness modeling in chronic obstructive pulmonary disease: a new patient-level simulation model suitable to evaluate stratified medicine. *Value Health*. 2019;22(3):313–21.

35. Buhl R, Maltais F, Abrahams R, et al. Tiotropium and olodaterol fixed-dose combination versus mono-components in COPD (GOLD 2-4). *Eur Respir J*. 2015;45:969–79.

36. Cooper CB, Celli BR, Jardim JR, et al. Treadmill endurance during 2-year treatment with tiotropium in patients with COPD: a randomized trial. *Chest*. 2013;144:490–7.

37. Tashkin DP, Celli B, Senn S, et al; UPLIFT Study Investigators. A 4-year trial of tiotropium in chronic obstructive pulmonary disease. *N Engl J Med*. 2008;359:1543–54.

38. Vogelmeier C, Hederer B, Glaab T, et al; POET-COPD Investigators. Tiotropium versus salmeterol for the prevention of exacerbations of COPD. *N Engl J Med*. 2011;364: 1093–103.

39. Wise RA, Anzueto A, Cotton D, et al; TIOSPIR Investigators. Tiotropium Respimat inhaler and the risk of death in COPD. *N Engl J Med*. 2013;369:1491–501.

40. R survival package documentation. Available from: https://cran.r-project.org/web/packages/survival

41. R lme4 package documentation. Available from: https://cran.r-project.org/web/packages/lme4

42. R nlme package documentation. Available from: https://cran.r-project.org/web/packages/nlme

43. R MASS package documentation. Available from: https://cran.r-project.org/web/packages/MASS

44. Briggs AH, Weinstein MC, Fenwick EAL, et al; ISPOR-SMDM Modeling Good Research Practices Task Force. Model parameter estimation and uncertainty: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-6. *Value Health*. 2012;15:835–42.

45. Groot Koerkamp B, Weinstein MC, Stijnen T, Heijenbrok-Kal MH, Hunink MG. Uncertainty and patient heterogeneity in medical decision models. *Med Decis Making*. 2010;30(2):194–205.

46. Vemer P, Goossens LM, Rutten-van Mölken MP. Not simply more of the same: distinguishing between patient heterogeneity and parameter uncertainty. *Med Decis Making*. 2014;34(8):1048–58.

47. Latimer N. NICE DSU Technical Support Document 14: Survival analysis for economic evaluations alongside clinical trials—extrapolation with patient-level data. 2013. Available from: http://nicedsu.org.uk/technical-support-documents/survival-analysis-tsd