VALUE IN HEALTH ▮ (2016) ▮▮▮–▮▮▮

Available online at www.sciencedirect.com

ScienceDirect

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

External Validation of Health Economic Decision Models for Chronic Obstructive Pulmonary Disease (COPD): Report of the Third COPD Modeling Meeting

Martine Hoogendoorn, PhD[1,], Talitha L. Feenstra, PhD[2,3], Yumi Asukai, MSc[4], Andrew H. Briggs, DPhil[5], Ryan N. Hansen, PharmD, PhD[6], Reiner Leidl, PhD[7], Nancy Risebrough, PhD[8], Yevgeniy Samyshkin, MSc[4], Margarethe Wacker, MSc[7], Maureen P.M.H. Rutten-van Mölken, PhD[1]*

[1]*Institute for Medical Technology Assessment (iMTA), Erasmus University Rotterdam, Rotterdam, The Netherlands;* [2]*Department for Prevention and Health Services Research, National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands;* [3]*Department of Epidemiology, University Medical Centre Groningen, Groningen, The Netherlands;* [4]*IMS Health, Economics and Outcomes Research and Real-World Evidence Solutions, London, UK;* [5]*Institute of Health and Wellbeing, University of Glasgow, Glasgow, UK;* [6]*Pharmaceutical Outcomes Research and Policy Program, School of Pharmacy, University of Washington, Seattle, WA, USA;* [7]*Helmholtz Zentrum München, Institute of Health Economics and Health Care Management, Member of the German Center for Lung Research, Comprehensive Pneumology Center Munich, Neuherberg, Germany;* [8]*ICON Health Economics, Toronto, Ontario, Canada*

ABSTRACT

Objectives: To validate outcomes of presently available chronic obstructive pulmonary disease (COPD) cost-effectiveness models against results of two large COPD trials—the 3-year TOwards a Revolution in COPD Health (TORCH) trial and the 4-year Understanding Potential Long-term Impacts on Function with Tiotropium (UPLIFT) trial. **Methods:** Participating COPD modeling groups simulated the outcomes for the placebo-treated groups of the TORCH and UPLIFT trials using baseline characteristics of the trial populations as input. Groups then simulated treatment effectiveness by using relative reductions in annual decline in lung function and exacerbation frequency observed in the most intensively treated group compared with placebo as input for the models. Main outcomes were (change in) total/severe exacerbations and mortality. Furthermore, the absolute differences in total exacerbations and quality-adjusted life-years (QALYs) were used to approximate the cost per exacerbation avoided and the cost per QALY gained. **Result:** Of the six participating models, three models reported higher total exacerbation rates than observed in the TORCH trial (1.13/patient-year) (models: 1.22–1.48). Four models reported higher rates than observed in the UPLIFT trial (0.85/patient-year) (models: 1.13–1.52). Two models reported higher mortality rates than in the TORCH trial (15.2%) (models: 20.0% and 30.6%) and the UPLIFT trial (16.3%) (models: 24.8% and 36.0%), whereas one model reported lower rates (9.8% and 12.1%, respectively). Simulation of treatment effectiveness showed that the absolute reduction in total exacerbations, the gain in QALYs, and the cost-effectiveness ratios did not differ from the trials, except for one model. **Conclusions:** Although most of the participating COPD cost-effectiveness models reported higher total exacerbation rates than observed in the trials, estimates of the absolute treatment effect and cost-effectiveness ratios do not seem different from the trials in most models.

Keywords: COPD, cost-effectiveness, external validation, model.

Introduction

Since 2004, several cost-effectiveness models for chronic obstructive pulmonary disease (COPD) have been developed and published [1–13]. Some of these models were specifically built to extrapolate single-trial results to a longer time horizon and support reimbursement decisions for newly developed drugs [4,6,8]. Other models used various data sources as input and are able to evaluate a wide range of different COPD interventions [3,7,10,12]. As a result of differences in data input, the models may refer to different populations of patients with COPD.

Because of their increasing role in decision making, it is very important that these cost-effectiveness models reflect the disease process and disease progression in COPD in an accurate way. Therefore, validation is a crucial part of model development [14]. One of the most important types of validation is external validation, which refers to comparing model outcomes against data from epidemiologic studies, clinical trials, or claims databases, preferably not used to build the model [14].

Since 2011, a worldwide network of researchers involved in COPD modeling (COPD modeling teams, pharmaceutical companies interested in COPD modeling, epidemiologists, clinicians, etc.)

* *Address correspondence to*: Martine Hoogendoorn, Institute for Medical Technology Assessment (iMTA), Erasmus University Rotterdam, P.O. Box 1738, Rotterdam 3000 DR, The Netherlands.

E-mail: hoogendoorn@imta.eur.nl.

1098-3015$36.00 – see front matter

http://dx.doi.org/10.1016/j.jval.2016.10.016

come together for 1-day annual meetings in Amsterdam to discuss and compare the currently available COPD models, collaborate, and share best practices about COPD modeling. During the second meeting in 2012, the models were cross-validated against each other to assess which differences in model structure, assumptions, and input data had the highest impact on the results of the models [15]. The main topic of the third meeting organized in 2014 was patient heterogeneity in COPD models [16]. Another topic of the third meeting was external validation of the models, which is the focus of this article.

The aim of the present article was to describe the validation of the outcomes of presently available COPD cost-effectiveness models against the results of two large clinical COPD trials and to assess the impact of the observed differences in outcomes on the cost-effectiveness ratio.

Methods

In the spring of 2014, modeling groups that participated in previous meetings as well as new groups were invited to participate in the modeling challenge for the third meeting. The challenge consisted of two components. For the first component, groups were requested to simulate outcomes for the placebo-treated groups of two large clinical COPD trials. For the second component, groups were asked to simulate the treatment effectiveness observed in the same trials. All results of the model simulations were reported in a structured format in Microsoft Excel and sent to the organizers of the meeting 2 weeks in advance. A summary of the combined results was circulated to all participants shortly before the meeting to give them the opportunity to reflect on the outcomes. During the meeting, results were presented and discussed to find possible explanations for deviations of the model outcomes from the trial results.

Clinical Trials

For this validation study, outcomes of two large long-term clinical trials in COPD were used: the TOwards a Revolution in COPD Health (TORCH) trial and the Understanding Potential Long-term Impacts on Function with Tiotropium (UPLIFT) trial [17,18]. In the 3-year TORCH trial, patients were randomly assigned to four treatment groups: 1) placebo, defined as all COPD medications except for long-acting bronchodilators (LABAs) and inhaled corticosteroids (ICSs); 2) salmeterol 50 μg; 3) fluticasone 500 μg; and 4) salmeterol 50 μg plus fluticasone 500 μg. The primary outcome of the trial was all-cause mortality. Secondary outcomes were exacerbations, health status, and lung function decline. The hazard ratio for mortality in the combination-therapy group compared with the placebo group was 0.825 (95% confidence interval [CI] 0.681–1.002) [16]. Compared with placebo the combination-therapy group had a significant reduction in exacerbations (relative reduction [RR] = 0.75; 95% CI 0.69–0.81) and in annual decline in forced expiratory volume in 1 second (FEV_1) (0.9% vs. 1.5% predicted/y; RR = 0.6) [17,19].

In the 4-year UPLIFT trial, patients with COPD were randomly assigned to the placebo group, which was defined as all regular respiratory medication except for inhaled anticholinergics or tiotropium 18 μg plus all regular respiratory medications except other inhaled anticholinergics. Primary outcomes of the trial were the pre- and postbronchodilator yearly rate of decline in FEV_1, whereas secondary outcomes were health-related quality of life, exacerbations, and mortality. No difference was observed between the two groups in the rate of decline in FEV_1 (post-bronchodilator: 40 vs. 42 ml/y; RR = 0.95). The tiotropium group had a lower number of exacerbations (RR = 0.86; 95% CI 0.81–0.91) and less mortality (hazard ratio = 0.87; 95% CI 0.87–0.99) compared with the placebo group [18].

Modeling Challenge

To simulate the outcomes of the two trials, the modeling groups populated their models with the baseline characteristics of the patients in the placebo groups in the trials. Models were adjusted, if possible, for percentage of males, mean age, percentage of present smokers, and mean FEV_1% predicted (or the distribution over the Global initiative for chronic Obstructive Lung Disease [GOLD] severity stages: moderate, severe, and very severe COPD) (Table 1). Other model parameters, such as disease progression, exacerbation probabilities, mortality, and utilities, were left unchanged.

The time horizons of the model simulations were equal to the treatment duration in the trials. Hence, modelers were asked to simulate the outcomes for the placebo group of the 3-year TORCH trial taking into account that patients did not receive LABA or ICS. Furthermore, outcomes for the placebo group of the 4-year UPLIFT trial were simulated taking into account that patients used all regular respiratory medication except other anticholinergics. Outcomes reported and compared with the trial results were total number of exacerbations per patient-year, total number of severe exacerbations per patient-year, and percentage of patients who died (means plus uncertainty intervals). Exacerbations in these analyses were defined as an increase in symptoms requiring treatment with antibiotics and/or systemic corticosteroids (moderate exacerbations) and/or hospitalization (severe exacerbation).

For the second component of the modeling challenge, the modeling groups were asked to simulate the relative treatment effectiveness as observed in the trials. Relative treatment effectiveness was defined as the RR in annual decline in lung function and exacerbations between the most intensively treated group (TORCH: salmeterol/fluticasone; UPLIFT: tiotropium) and the placebo group. The observed RRs in annual decline in lung function and exacerbations were applied to the model input values of these parameters used to simulate the outcomes for the placebo group. This method is regarded as an appropriate

Table 1 – Baseline characteristics of patients in the placebo groups of the TORCH and UPLIFT trials used as starting population of the model simulations [17,18].

Trial	TORCH placebo	UPLIFT placebo
N	1524	3006
Males	76%	74%
Age (y), mean ± SD	65 ± 8	65 ± 9
Current smokers	43%	30%
Post-FEV_1% predicted, mean ± SD	44 ± 12	47 ± 13
Severity distribution		
GOLD II: moderate COPD	35%	45%
GOLD III: severe COPD	50%	44%
GOLD IV: very severe COPD	15%	9%

FEV_1% predicted, forced expiratory volume in 1 s as percentage of the predicted value; GOLD, Global initiative for chronic Obstructive Lung Disease; TORCH, TOwards a Revolution in COPD Health; UPLIFT, Understanding Potential Long-term Impacts on Function with Tiotropium.

way of modeling treatment effectiveness in cost-effectiveness models [20]. For the present study, the RRs were applied in two steps. First, the RR in annual lung function decline was applied and then the RR in exacerbations was added on top of that. The absolute differences in total exacerbations, severe exacerbations, and mortality between treatment options were calculated and compared with the observed differences in the trials. On the basis of the 95% uncertainty intervals around the mean model outcomes, the standard errors (SEs) for the model outcomes were calculated as the upper limit of the uncertainty interval minus the lower limit divided by 3.92. These SEs were combined with the mean outcomes and SEs observed in the two COPD trials to calculate the mean differences between the model and trial outcomes. The mean difference was calculated as the mean model outcome minus the mean trial outcome. The SE around this difference was calculated as the square root of the sum of the quadratic SE of the model outcomes and the quadratic SE of the trial outcomes and used to calculate the 95% CI around the mean difference to assess whether the difference was significant.

In addition, the incremental cost-effectiveness ratios (ICERs) of treatment with the combination of salmeterol/fluticasone and treatment with tiotropium compared with placebo were approximated using the same unit costs in all the models and the trials. The ICER was estimated as the additional medication costs in the treated group as compared with the placebo group minus the savings in exacerbation-related costs divided by the difference in effect, the number of exacerbations avoided or the number of quality-adjusted life-years (QALYs) gained. Savings in maintenance costs could not be included, because these costs could not be calculated on the basis of the information in the publications of the trials.

Overview of Participating Models

Six COPD cost-effectiveness models participated in the external validation challenge: the indacaterol COPD model represented by Asukai [8], the Galaxy COPD model represented by Briggs [13,21,22], the US Dynamic Cohort COPD model represented by Hansen [12], the Dutch Dynamic population COPD progression model represented by Hoogendoorn [3,23,24], the roflumilast COPD model represented by Samyshkin [9,25,26], and the German comprehensive care COPD model represented by Wacker [10]. Five of the six participating models were so-called state-transition models, with states representing COPD severity on the basis of GOLD lung function classification: mild, moderate, severe, and very severe COPD [27]. As a result, the FEV_1% predicted was the key parameter in these models, and input data for disease progression, exacerbation frequency, mortality, quality of life, and costs were specified by GOLD stage. The models simulated the progression of disease over time to worse COPD stages and the occurrence of exacerbations and death. The Galaxy COPD model represented by Briggs was most recently developed and it used a different modeling approach. In this model, causal relationships between disease parameters such as lung function, exacerbations, symptoms, and exercise capacity, and outcomes such as mortality, quality of life, and costs were estimated and all these equations were linked together taking into account changes over time [13,21,22]. All models had a lifetime horizon. Appendix Table 1 in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2016.10.016 includes more details on the model structure, patient population, and input data for the different models. A short description of the participating models except for the model of Briggs has been published elsewhere [15], whereas extensive details can be found in the individual publications [3,8–10,12,13].

Results

Exacerbations

Five of the six participating models included exacerbation rates as outcome. The model of Hansen reported exacerbation days instead of rates and was therefore not included in the validation of exacerbations. Results of the model simulations for exacerbations in the placebo arms of the TORCH and UPLIFT trials are shown in Figures 1 and 2, respectively.

Three of the five models (Hoogendoorn, Samyshkin, and Wacker) reported a higher total exacerbation rate than the rate observed in the TORCH trial (1.13 per patient-year [17]) (Fig. 1). For the comparison with the UPLIFT trial, four models (Briggs, Hoogendoorn, Samyshkin, and Wacker) reported higher total exacerbation rates compared with the trial (0.85 per patient-year [18]) (Fig. 1). The level of uncertainty around simulated exacerbation rates varied substantially between the models.

Results for severe exacerbations (Fig. 2) showed that for both the TORCH and the UPLIFT trials the model of Briggs reported higher severe exacerbation rates than observed in the trials, whereas rates were lower in the model of Asukai. The model of Wacker estimated a lower severe exacerbation rate than observed in the TORCH trial, and the model of Samyshkin reported a higher severe exacerbation rate than observed in the UPLIFT trial.

Mortality

Results for mortality are shown in Figure 3. The 3-year mortality rate in the placebo group of the TORCH trial was 15.2% [17]. The mortality rate in the placebo group of the UPLIFT trial was 16.3% over a 4-year period [18]. The model of Hansen reported lower mortality rates than both trials, whereas the two models of Hoogendoorn and Wacker reported higher rates. The model of Briggs estimated a lower mortality rate compared with the TORCH trial.

Treatment Effectiveness

Table 2 presents the changes in total exacerbations, severe exacerbations, and mortality after simulation of treatment effectiveness observed in the TORCH and UPLIFT trials. Applying the RR for annual decline in lung function did not have much impact on the results. None of the exacerbation and mortality outcomes

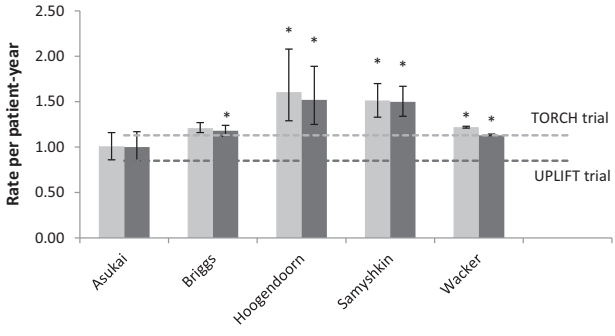

Fig. 1 – Comparison of model simulations for total exacerbations with the trial results for the TORCH trial (placebo group), 3-y time horizon (gray), and the UPLIFT trial (placebo group), 4-y time horizon (black), including 95% uncertainty intervals. TORCH, TOwards a Revolution in COPD Health; UPLIFT, Understanding Potential Long-term Impacts on Function with Tiotropium; *, different from the trial result.

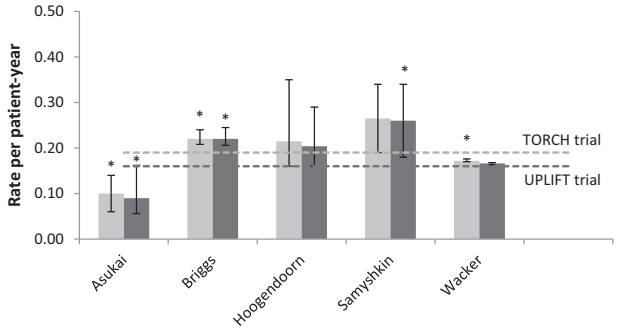

Fig. 2 – Comparison of model simulations for severe exacerbations with the trial results for the TORCH trial (placebo group), 3-y time horizon (gray), and the UPLIFT trial (placebo group), 4-y time horizon (black), including 95% uncertainty intervals. TORCH, TOwards a Revolution in COPD Health; UPLIFT, Understanding Potential Long-term Impacts on Function with Tiotropium; *, different from the trial result.

changed compared with placebo. Applying both the RR for annual decline and the RR for exacerbations resulted in changes in total exacerbations and severe exacerbations compared with placebo. For the TORCH trial, four models (Asukai, Briggs, Samyshkin, and Wacker) reported a significantly lower total exacerbation rate in the most intensively treated group compared with the placebo group. Severe exacerbations were reduced in two models (Briggs and Wacker). For the UPLIFT trial, there was a reduction in total and severe exacerbations only in the model of Wacker. Mortality did not change in any of the models.

Cost-Effectiveness

For the TORCH trial, simulation of treatment effectiveness showed that the absolute reduction in total exacerbations between the salmeterol/fluticasone and the placebo group estimated by the models ranged from −0.22 (95% uncertainty interval [UI] −0.44 to −0.003) for the model of Asukai to −0.44 (95% UI −0.45 to 0.43) for the model of Wacker compared with −0.28 (95% UI −0.36 to −0.20) observed in the trial (Table 3). Only for the model of Wacker, the absolute difference in total exacerbations was higher than the trial. The estimated cost per exacerbation avoided varied between €1100 and €2600. Because of the high uncertainty around the number of exacerbations avoided in most

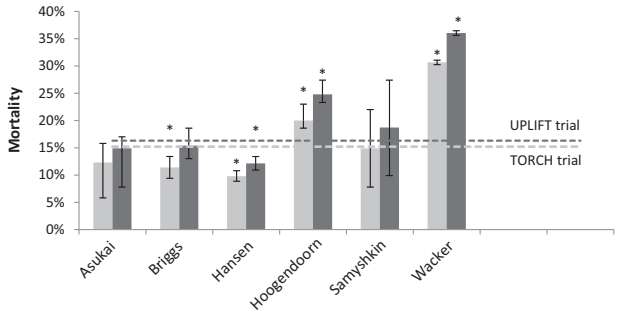

Fig. 3 – Comparison of model simulations for mortality with the trial results for the TORCH trial (placebo group), 3-y time horizon (gray), and the UPLIFT trial (placebo group), 4-y time horizon (black), including 95% uncertainty intervals. TORCH, TOwards a Revolution in COPD Health; UPLIFT, Understanding Potential Long-term Impacts on Function with Tiotropium; *, different from the trial result.

models, most ratios seemed comparable with the trial estimate of €2000 per exacerbation avoided. The estimated gain in QALYs varied between 0.003 (95% UI −1.66 to 1.66) for the model of Asukai and 0.033 (95% UI −0.073 to 0.80) for the model of Samyshkin compared with 0.081 (95% CI 0.028 to 0.134) observed in the TORCH trial [28]. The model of Wacker reported a lower number of QALYs gained compared with the trial, resulting in a higher cost per QALY gained ratio compared with the trial.

For the UPLIFT trial, the model estimates for the absolute reduction in total exacerbations between tiotropium and placebo ranged from −0.11 (95% UI −0.33 to 0.11) for the model of Asukai to −0.21 (95% UI −0.44 to 0.07) for the model of Samyshkin compared with −0.12 (95% CI −0.18 to −0.07) observed in the trial. For none of the models, the absolute reduction in exacerbations was different compared with the trial.

Discussion

This study aimed to validate the outcomes of presently available cost-effectiveness models in COPD against published results of two large clinical trials to show how well the models predict exacerbations and mortality and how useful the models are to support decision making.

Results showed that most models tend to overestimate total exacerbations over the duration of the trial: three out of five models reported a higher total exacerbation rate than observed in the TORCH trial, whereas this was four out of five for the UPLIFT trial. These deviations might be influenced by differences in model structure or in parameter input (see Appendix Table 1 in Supplemental Materials). Some models used a meta-analysis as a basis for exacerbation risk (Hoogendoorn), and others used individual studies that were partly included in the meta-analysis (Wacker) or used exacerbation rates obtained from specific clinical trials (Asukai and Samyshkin). Overestimation of the baseline exacerbation risk in a COPD cost-effectiveness model could result in an overestimation of the absolute treatment effect (i.e., the number of exacerbations avoided) and therefore in an underestimation of the cost-effectiveness ratio (i.e., the cost per exacerbation avoided). Therefore, the second part of the present study focused on the estimation of treatment effects and the impact of differences on the cost-effectiveness. Simulation of treatment effectiveness resulted in reductions in total exacerbations compared with placebo in four models for the TORCH trial and in two models for the UPLIFT trial. Given the large uncertainty around the model outcomes, in most models the absolute reduction in exacerbations was not different from the value observed in the trial, resulting in comparable estimates of the cost per exacerbation avoided.

All-cause mortality over the duration of the trial was underestimated by one model and overestimated by two models. Again, differences in parameter input are important for the deviations and have been reported elsewhere [15]. Simulation of treatment effectiveness in terms of a reduction in annual decline and exacerbations reduced mortality in all models. This decrease was, however, small and not significant in any of the models. The reductions in mortality observed in the trials were (borderline) significant and seemed higher than the model estimates, indicating that there probably is an additional independent effect of treatment on mortality that was not related to the reduction in annual decline in lung function and the reduction in exacerbations. The presently observed reduction in mortality in the models is solely the result of the decrease in lung function decline and severe exacerbations, because no independent effect on mortality is included in the model estimates. As a result, the number of QALYs gained reported by the models appeared also lower than the trial, although for only one model the gain in

Table 2 – Model outcomes for the most intensively treated group compared with placebo for total exacerbations, severe exacerbations, and mortality after simulating treatment effectiveness as observed in the trials*.

Analysis	Absolute difference compared with placebo						
	Trial	Model outcomes					
		Asukai	Briggs	Hansen	Hoogendoorn	Samyshkin	Wacker
TORCH trial							
Placebo + RR_annual decline							
Total exacerbations	NA	0	−0.020	NA	−0.017	−0.017	0.010
Severe exacerbations	NA	0	−0.01	NA	−0.002	−0.011	0.001
Mortality	NA	−0.1%	−0.5%	0%	−0.1%	−0.6%	0.1%
Placebo + RR_annual decline + RR_exacerbations							
Total exacerbations	−0.28†	−0.222†	−0.320†	NA	−0.416	−0.391†	−0.436†
Severe exacerbations	−0.03†	−0.017	−0.040†	NA	−0.038	−0.054	−0.028†
Mortality	−2.6%	−0.1%	−0.7%	−0.04%	−1.4%	−1.5%	−0.5%
UPLIFT trial							
Placebo + RR_annual decline							
Total exacerbations	NA	0	−0.030	NA	0.004	−0.002	−0.001
Severe exacerbations	NA	−0.075	−0.012	NA	0.001	−0.029	−0.001
Mortality	NA	0%	−0.6%	0%	0%	−0.8%	0%
Placebo + RR_annual decline + RR_exacerbations							
Total exacerbations	−0.12†	−0.114	−0.190†	NA	−0.209	−0.212	−0.138†
Severe exacerbations	−0.01	−0.074	−0.025	NA	−0.1012	−0.044	−0.010†
Mortality	−1.9%†	0%	−0.8%	−0.01%	−0.5%	−1.2%	−0.3%†

NA, not available; RR, relative reduction; TORCH, TOwards a Revolution in COPD Health; UPLIFT, Understanding Potential Long-term Impacts on Function with Tiotropium.

* Treatment effectiveness TORCH salmeterol/fluticasone vs. placebo: annual decline in lung function RR = 0.6, total exacerbations RR = 0.75, severe exacerbations RR = 0.83 [17,19]. Treatment effectiveness UPLIFT tiotropium vs. placebo: annual decline in lung function RR = 0.95, total exacerbations RR = 0.86, severe exacerbations RR = 0.94 [18].

† Different from placebo.

QALYs was significantly lower than the trial, resulting in a substantial higher cost per QALY gained. Correct implementation of full treatment effectiveness would have required calibration of the models by adjusting all-cause mortality rates further to the effect as observed in the trials. The underestimation of the mortality effect mainly affected the cost per QALY gained, but did not have much impact on the cost per exacerbation avoided.

Results of this study also showed that the level of uncertainty around the simulated outcomes varied substantially between models. The model of Wacker reported very little uncertainty around the estimated exacerbation and mortality rates, whereas the models of Hoogendoorn and Samyshkin reported wide uncertainty intervals around the outcomes. The reported uncertainty was used to calculate whether results were different from the trial and whether simulated treatment outcomes changed. As expected, models with a high level of uncertainty were less likely to find significant differences compared with the trial than were models with little uncertainty. The same was true for treatment effectiveness. A similar absolute reduction in exacerbations in two models could be significant in one model, whereas it was not in the other model because of the difference in uncertainty around the outcomes.

The present study focused on the outcomes exacerbations and mortality. Annual decline in lung function would have been an interesting parameter for validation as well, but it was not possible to use this parameter for validation because decline was modeled differently in all models. Some models included annual decline in FEV_1 in liters, and other models included annual decline in FEV_1% predicted. Most models, however, included transition rates between COPD severity stages, which were calculated on the basis of annual decline in lung function. Especially in these models annual lung function decline was not a model outcome. Results, however, showed that the impact of annual decline in lung function on the outcomes was limited. Applying an RR as large as 0.6 for annual lung function decline hardly impacted the exacerbations and mortality over a 3-year time horizon (Table 2).

The main challenge in simulating the TORCH and UPLIFT trials was accounting for COPD treatment. Modeling groups were asked to simulate the placebo arm of the TORCH trial taking into account that patients did not receive LABA or ICS. For the placebo arm of the UPLIFT trial, groups were asked to take into account that patients received all regular respiratory medications except other anticholinergics. For most models, however, the default baseline input parameters were considered to reflect placebo. For the models of Asukai, Hansen, Hoogendoorn, and Wacker, most of the input data were based on data sources assuming no treatment or minimal treatment, that is, short-acting bronchodilators. So for these models, results using the default baseline input parameters were best comparable with the TORCH placebo arm and less comparable with the UPLIFT placebo-treated group. This was, however, not reflected in the results. In the models of Hoogendoorn and Wacker, exacerbations and mortality were overestimated for both the UPLIFT trial and the TORCH trial, and the model of Asukai overestimated severe exacerbations for both trials. The newest model of Briggs was completely based on the ECLIPSE data in which ICS, LABA, and tiotropium were frequently used [29]. Results of this model were therefore best

Table 3 – Absolute difference in effect between the most intensively treated group and placebo and the approximated cost-effectiveness ratios assuming standardized unit costs across models, mean (95% uncertainty intervals)*.

	TORCH trial: salmeterol/fluticasone vs. placebo				UPLIFT trial: tiotropium vs. placebo	
	Absolute difference in total exacerbations per patient-year	Cost per exacerbation avoided (€)	Absolute difference in QALYs	Cost per QALY gained (€)	Absolute difference in total exacerbations per patient-year	Cost per exacerbation avoided (€)
Trial	−0.28 (−0.36 to −0.20)	2,000	0.081 (0.028 to 0.13)	18,200	−0.12 (−0.18 to −0.07)	4,900
Asukai	−0.22 (−0.44 to −0.003)	2,600	0.003 (−1.66 to 1.66)	515,200	−0.11 (−0.33 to 0.11)	5,000
Briggs	−0.32 (−0.40 to −0.24)	1,400	0.024 (−0.01 to 0.060)	60,500	−0.19 (−0.27 to −0.11)	2,600
Hansen	NA	NA	0.025 (−0.017 to 0.066)	NA	NA	NA
Hoogendoorn	−0.42 (−0.89 to 0.06)	1,200	0.029 (−0.17 to 0.24)	43,800	−0.21 (−0.63 to 0.22)	2,600
Samyshkin	−0.39 (−0.63 to −0.15)	1,100	0.033 (−0.73 to 0.80)	35,000	−0.21 (−0.44 to 0.07)	2,000
Wacker	−0.44 (−0.45 to −0.43)†	1,200	0.012 (0.004 to 0.020)†	107,600	−0.14 (−0.15 to −0.13)	3,900

NA, not available; QALY, quality-adjusted life-year; TORCH, TOwards a Revolution in COPD Health; UPLIFT, Understanding Potential Long-term Impacts on Function with Tiotropium.

* Only medication costs and costs for treating exacerbations were included in the cost calculation. Costs were calculated assuming the following unit costs: salmeterol/fluticasone, €1.80/d; tiotropium, €1.60/d; moderate exacerbation, €100; severe exacerbation, €4,000.

† Different compared with the trial.

comparable with the UPLIFT trial and should underestimate outcomes for the TORCH placebo group. Results showed that this was true for mortality, but not for exacerbations. Adjustment of the baseline model values for a different treatment or treatment mix is very difficult in most models because it requires relative risks for a treatment or treatment mix for different parameters compared with the default baseline value used in the model, which are often not available.

A basic issue in validation models with trial data is that models reflect disease progression and mortality of specific COPD populations. Populations referred to are key features of models. If results are intended to be extrapolated for a trial population, trial results are the appropriate criterion standard for validation. Some models in this study yet refer to nontrial populations (Hansen, Hoogendoorn, and Wacker) and used, for example, mortality data of the general population. Results of this study show how well the included models refer to the two trial populations and would be suitable for respective extrapolation.

A limitation of the present study was that we could only approximate the cost-effectiveness ratios. For the TORCH trial, a cost-effectiveness study has been published reporting the cost per QALY gained [28]. This study, however, included several types of costs that are not included in the models, such as non–COPD-related medications and non–COPD-related hospitalizations and did not present unit costs used because it was a multinational trial and unit costs are country-specific. Therefore, we approximated the cost-effectiveness ratios by including only the medication costs provided as treatment in the trial and the exacerbation-related costs using standardized unit costs. Costs for maintenance treatment were not included because the information in the publications of the trials was too limited to calculate this type of cost. Omission of this type of cost likely had minimal bias on the outcomes; for example, in the model of Hoogendoorn, the difference in the cost per exacerbation avoided including or excluding maintenance costs was less than €10.

Estimates of the cost-effectiveness ratios were dependent on the standardized unit costs used to calculate costs. Changing the unit costs, however, showed that the relative difference in the ratios compared with the trial was rather constant.

Appropriate data sources for external validation should include detailed description of the patient population, setting, possible treatment options, and outcomes and should preferably have a large size [14]. Therefore, clinical trials are a common source for this type of validation. Detailed information is needed to adjust the model as much as possible to the data source used. In the present study, the baseline patient population of the models was adjusted for percentage of males, mean age, percentage of current smokers, and mean FEV_1% predicted. These patient characteristics were chosen because most models included these characteristics. The model of Briggs, however, included several other patient characteristics, such as body mass index, comorbidities, previous exacerbations, modified Medical Research Council dyspnea scale (mMRC), and 6-minute walk test distance. If this model would have been adjusted for these parameters as well, results may have been different.

Databases used for external validation should preferably not have been used to build the model. Some of the participating models, however, used the UPLIFT and TORCH trials to estimate input parameters. The model of Asukai used the lung function decline reported in the UPLIFT trial to estimate transition rates between COPD severity stages. The model of Wacker based the lung function decline in severe COPD on the results of the TORCH trial, and the exacerbation frequency by GOLD stage in the model of Hoogendoorn was estimated on the basis of a meta-analysis of 19 studies including the TORCH trial. Although this is a limitation of the study, none of the models is completely based on

the UPLIFT or TORCH trial. Furthermore, the availability of well-performed COPD trials with a follow-up of several years that could be used for this validation study was limited.

Conclusions

This validation study showed that most of the six COPD models that participated in this study overestimated the total exacerbation frequency when compared with two large clinical trials. Mortality was structurally underestimated by one model and overestimated by two models. Simulation of treatment effectiveness resulted in comparable differences in the absolute number of exacerbations avoided and QALYs gained between the models and the trials because most models reported much uncertainty around the outcomes. As a result of this large uncertainty, estimates of the cost per exacerbation avoided and the cost per QALY gained also do not seem different from the trials in most models.

Source of financial support: This study was financially supported by Boehringer Ingelheim International (Germany), GlaxoSmithKline (the Netherlands), Novartis International (Switzerland), and Takeda Pharmaceuticals International (Switzerland).

Supplemental Materials

Supplemental material accompanying this article can be found in the online version as a hyperlink at http://dx.doi.org/10.1016/j.jval.2016.10.016 or, if a hard copy of article, at www.valueinhealthjournal.com/issues (select volume, issue, and article).

REFERENCES

[1] Borg S, Ericsson A, Wedzicha J, et al. A computer simulation model of the natural history and economic impact of chronic obstructive pulmonary disease. Value Health 2004;7:153–67.
[2] Spencer M, Briggs AH, Grossman RF, Rance L. Development of an economic model to assess the cost effectiveness of treatment interventions for chronic obstructive pulmonary disease. Pharmacoeconomics 2005;23:619–37.
[3] Hoogendoorn M, Rutten-van Molken MP, Hoogenveen RT, et al. Developing and applying a stochastic dynamic population model for chronic obstructive pulmonary disease. Value Health 2011;14:1039–47.
[4] Rutten-van Molken MP, Oostenbrink JB, Miravitlles M, Monz BU. Modeling the 5-year cost effectiveness of tiotropium, salmeterol and ipratropium for the treatment of chronic obstructive pulmonary disease in Spain. Eur J Health Econ 2007;8:123–35.
[5] Dal NR, Eandi M, Pradelli L, Iannazzo S. Cost-effectiveness and healthcare budget impact in Italy of inhaled corticosteroids and bronchodilators for severe and very severe COPD patients. Int J Chron Obstruct Pulmon Dis 2007;2:169–76.
[6] Earnshaw SR, Wilson MR, Dalal AA, et al. Cost-effectiveness of fluticasone propionate/salmeterol (500/50 microg) in the treatment of COPD. Respir Med 2009;103:12–21.
[7] Nielsen R, Johannessen A, Benediktsdottir B, et al. Present and future costs of COPD in Iceland and Norway: results from the BOLD study. Eur Respir J 2009;34:850–7.
[8] Price D, Gray A, Gale R, et al. Cost-utility analysis of indacaterol in Germany: a once-daily maintenance bronchodilator for patients with COPD. Respir Med 2011;105:1635–47.
[9] Samyshkin Y, Kotchie RW, Mork AC, et al. Cost-effectiveness of roflumilast as an add-on treatment to long-acting bronchodilators in the treatment of COPD associated with chronic bronchitis in the United Kingdom. Eur J Health Econ 2014;15:69–82.
[10] Menn P, Leidl R, Holle R. A lifetime Markov model for the economic evaluation of chronic obstructive pulmonary disease. Pharmacoeconomics 2012;30:825–40.
[11] Asukai Y, Baldwin M, Fonseca T, et al. Improving clinical reality in chronic obstructive pulmonary disease economic modeling: development and validation of a micro-simulation approach. Pharmacoeconomics 2013;31:151–61.
[12] Hansen RN, Xu X, Sullivan SD. A dynamic cohort model of chronic obstructive pulmonary disease and its treatments. Value Health 2012;15:A54.
[13] Briggs A, Baker T, Risebrough NA, et al. Development of the Galaxy chronic obstructive pulmonary disease (COPD) model using data from ECLIPSE: internal validation of a linked-equations cohort model. Med Decis Making (published online ahead of print June 17, 2016). pii: 0272989X16653118.
[14] Eddy D, Hollingworth W, Caro JJ, et al. Model transparancy and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group—part 4. Value Health 2012;15:843–50.
[15] Hoogendoorn M, Feenstra TL, Asukai Y, et al. Cost-effectiveness models for chronic obstructive pulmonary disease: cross-model comparison of hypothetical treatment scenarios. Value Health 2014;17:525–36.
[16] Hoogendoorn M, Feenstra T, Asukai Y, et al. Patient heterogeneity in health economic decision models for chronic obstructive pulmonary disease: are current models suitable to evaluate personalized medicine? Value Health 2016;19:800–10.
[17] Calverley PM, Anderson JA, Celli B, et al. Salmeterol and fluticasone propionate and survival in chronic obstructive pulmonary disease. N Engl J Med 2007;356:775–89.
[18] Tashkin DP, Celli B, Senn S, et al. A 4-year trial of tiotropium in chronic obstructive pulmonary disease. N Engl J Med 2008;359:1543–54.
[19] Celli BR, Thomas NE, Anderson JA, et al. Effect of pharmacotherapy on rate of decline of lung function in chronic obstructive pulmonary disease: results from the TORCH study. Am J Respir Crit Care Med 2008;178:332–8.
[20] Weinstein MC, O'Brien B, Hornberger J, 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 Health 2003;6:9–17.
[21] Exuzides A, Colby C, Briggs A, et al. Statistical modeling of disease progression for chronic obstructive pulmonary disease using data from the ECLIPSE study. Med Decis Making (published online ahead of print October 8, 2015). pii: 0272989X15610781.
[22] Risebrough NA, Briggs A, Baker TM, et al. Validating a model to predict disease progression outcomes in patients with COPD. Value Health 2014;7:A560–1.
[23] Hoogendoorn M, Rutten-van Molken MP, Hoogenveen RT, et al. A dynamic population model of disease progression in COPD. Eur Respir J 2005;26:223–33.
[24] Hoogendoorn M, Feenstra TL, Hoogenveen RT, Rutten-van Molken MP. Long-term effectiveness and cost-effectiveness of smoking cessation interventions in patients with COPD. Thorax 2010;65:711–8.
[25] Hertel N, Kotchie RW, Samyshkin Y, et al. Cost-effectiveness of available treatment options for patients suffering from severe COPD in the UK: a fully incremental analysis. Int J Chron Obstruct Pulmon Dis 2012;7:183–99.
[26] Samyshkin Y, Schlunegger M, Haefliger S, et al. Cost-effectiveness of roflumilast in combination with bronchodilator therapies in patients with severe and very severe COPD in Switzerland. Int J Chron Obstruct Pulmon Dis 2013;8:79–87.
[27] Global Initiative for Chronic Obstructive Lung Disease. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. 2013. Available from: http://www.goldcopd.com. [Accessed June 15, 2013].
[28] Briggs AH, Glick HA, Lozano-Ortega G, et al. Is treatment with ICS and LABA cost-effective for COPD? Multinational economic analysis of the TORCH study. Eur Respir J 2010;35:532–9.
[29] Vestbo J, Edwards LD, Scanlon PD, et al. Changes in forced expiratory volume in 1 second over time in COPD. N Engl J Med 2011;365:1184–92.