University of Kentucky

From the Selected Works of David M. Mannino

December, 2008

Obstructive Lung Disease Models: What is Valid

Jill M. Ferdinands
David M. Mannino

Available at: https://works.bepress.com/david_mannino/59/

CLINICAL REVIEW

Obstructive Lung Disease Models: What Is Valid?

Jill M. Ferdinands¹ (zdn5@cdc.gov) and David M. Mannino² (Dmannino@uky.edu)

¹Influenza Division, Centers for Disease Control and Prevention, Atlanta, Georgia, USA

²Division of Pulmonary and Critical Care Medicine, University of Kentucky College of Medicine, Lexington, Kentucky, USA

ABSTRACT

Use of disease simulation models has led to scrutiny of model methods and demand for evidence that models credibly simulate health outcomes. We sought to describe recent obstructive lung disease simulation models and their validation. Medline and EMBASE were used to identify obstructive lung disease simulation models published from January 2000 to June 2006. Publications were reviewed to assess model attributes and four types of validation: first-order (verification/debugging), second-order (comparison with studies used in model development), third-order (comparison with studies not used in model development), and predictive validity. Six asthma and seven chronic obstructive pulmonary disease models were identified. Seven (54%) models included second-order validation, typically by comparing observed outcomes to simulations of source study cohorts. Seven (54%) models included third-order validation, in which modeled outcomes were usually compared qualitatively for agreement with studies independent of the model. Validation endpoints included disease prevalence, exacerbation, and all-cause mortality. Validation was typically described as acceptable, despite near-universal absence of criteria for judging adequacy of validation. Although over half of recent obstructive lung disease simulation models report validation, inconsistencies in validation methods and lack of detailed reporting make assessing adequacy of validation difficult. For simulation modeling to be accepted as a tool for evaluating clinical and public health programs, models must be validated to credibly simulate health outcomes of interest. Defining the required level of validation and providing guidance for quantitative assessment and reporting of validation are important future steps in promoting simulation models as practical decision tools.

INTRODUCTION

Asthma and chronic obstructive pulmonary disease, collectively known as obstructive lung disease, affect more than 30 million Americans (1–4). Understanding the natural history of these diseases and developing and implementing policies to reduce the morbidity and mortality from them, are important health goals (5). Disease simulation models—quantitative rep-

resentations of disease progression and factors that influence it—are often used to compare competing health interventions and explore their effectiveness and cost-effectiveness in situations where randomized or observational studies are logistically impossible, ethically questionable, or too resource-intensive to conduct. In an era when health care resources to address obstructive lung disease are shrinking and demands to demonstrate quantitative health impact of programs aimed at those diseases are increasing, there is escalating interest in the application of disease simulation models to forecast disease burden and explore effectiveness of alternative interventions. With increasing interest in disease simulation models as exploratory policy tools comes concomitant interest in the quality of these tools, especially in regard to establishing that they do a credible job of what they purport to do: quantitatively predict health outcomes under various scenarios. That is, there is a desire to demonstrate—insofar as is necessary—that the models are valid.

The value of disease simulation models is often diminished by practical skepticism that takes the form of such questions as,

Keywords: Chronic obstructive lung disease, Asthma, Natural history, Model, Validation
No funding was received for conduct of this study and generation of this manuscript.
Correspondence to:
Jill Ferdinands
Centers for Disease Control and Prevention
1600 Clifton Road, MS A-32
Atlanta, GA 30333, USA
email: zdn5@cdc.gov

"How do I know the model is right?" and "How do I know I can trust the model results?" Addressing these concerns in a consistent, logical manner is essential to the effective use of these tools and their acceptance by decision makers. Various issues related to these validity concerns have been addressed in recent guidance on best practices in use of modeling for health policy decisions (6–8). While these efforts to address model validation are helpful steps, additional guidance describing model validation standards would greatly assist clinical and public health policymakers in responding to important questions about model validity and, in so doing, enhance the acceptance of disease simulation models and the insights they offer.

To improve understanding of obstructive lung disease models in general and the methods used to validate them in particular, we sought to describe obstructive lung disease models reported in the recent peer-reviewed literature and characterize methods used to validate them. Our goal was to summarize the current "state of the art" for validation of obstructive lung disease simulation models and to recommend actions that may help to improve model validation methods, thus enhancing the usefulness of these tools.

MATERIALS AND METHODS

We define an obstructive lung disease simulation model as an analytic methodology (i.e., a sequence of logical mathematical computations) that links together evidence on obstructive lung disease from many sources to generate estimates of asthma or COPD disease events. In this sense, an obstructive lung disease simulation model is essentially a mechanistic representation of disease progression. An obstructive lung disease simulation model should be distinguished from the broader application that makes use of it, such as an economic evaluation. We have intentionally limited the scope of our assessment to examining the validity of disease simulation models themselves rather than the quality of economic evaluations and other applications that use them.

To identify recent obstructive lung disease simulation models, we conducted one primary and three secondary searches using PubMed and EMBASE databases. Details of the searches and search terms are given in the Supplement. Each search was limited to papers published in English between January 2000 and June 2006. All abstracts retrieved by these searches were hand-screened for eligibility, with particular emphasis on the Methods section of the abstract. Eligible abstracts included those that (a) reported on asthma or COPD (or synonyms such as *emphysema*) as the primary health outcomes of interest; and (b) appeared to describe use or development of a mathematical disease simulation model, as indicated by descriptors such as *Markov model*, *state-transition model*, *prediction model*, *simulation model*, *disease simulation*, *disease progression model*, *mathematical model*, *dynamic population model*, *decision analytic model*, *life table model*, and *natural history model*. For all abstracts that met screening criteria, we retrieved and reviewed the corresponding full publication for potential inclusion in this review.

If it was unclear if an abstract met screening criteria, the full paper was reviewed. Criteria for inclusion in this review included (a) publication between January 2000 and June 2006; (b) written in English; and (c) reported on use or development of a disease simulation model of asthma or COPD. We included as disease simulation models both state-transition models and dynamic population models. We excluded neural networks and simple decision tree models. If it could not be determined from the full paper if the model was indeed a disease simulation model, the paper was excluded, as were publications describing models still under development [e.g., the COPD model of Buist et al. (9, 10)].

Of 580 English-language abstracts reviewed, 13 reported use of an obstructive lung disease simulation model. Two of these were enhanced versions of previously-reported models. Because the enhanced models incorporated new model structures that could affect model validation, they were included as separate models.

Publications describing the models were reviewed to abstract information about model characteristics, including model purpose, type (e.g., state-transition or dynamic population model), attributes (e.g., number of disease states, cycle length, and time horizon), health outcomes modeled, and funding source. In addition to reviewing papers identified by the search strategy, we reviewed supplemental material about the models (e.g., technical reports) only if such information were publicly available and cited by the main papers. Because we were interested in assessing model validation using essentially the same body of information readily available to a typical audience of policy-makers, we intentionally made no attempt to obtain unpublished material or information available solely by personal communication with the authors.

Current terminology for describing validation methods is inconsistent, making categorization of validation methods challenging. Using an adaptation of terminology first proposed by David Eddy (11), we defined four types of model validation for the purposes of this study: first-order validity (verification and debugging), second-order validity (comparison of model estimates with published data from studies used to develop the model), third-order validity (comparison of model estimates with published data from studies not used to develop the model), and predictive validity. We recognize that a diverse set of terms has been used to describe these processes, including *calibration*. Because the distinction between calibration and validation is tenuous and seems to depend on whether the comparison of modeled and observed values is used to re-specify the model (in which case this feedback into model development is often deemed *calibration* rather than *validation*), we forego the term *calibration* in favor of specific definitions of validation. Similarly, because there is inconsistent use of the terms *internal validation* and *external validation* to refer to various validation techniques, we eliminated those descriptors in favor of the explicit definitions here.

We defined first-order validity as the process by which model structure is examined for coding errors, logical inconsistencies, and accuracy of mathematical calculations. Sometimes referred

to as debugging or model verification, this often includes logical checks on model programming conducted using null values, extreme values, and hypothetical conditions, in addition to the overall assessment of face validity proposed by Eddy as a test of first-order validity. First-order validation need not involve direct comparison of model predictions to other data sources. We defined second-order validation as comparison of model predictions *with existing data from studies that were used to parameterize, construct, or develop the model*. This includes, but is not limited to, comparison of model predictions to results of randomized trials used to develop subsets of model parameters.

We defined third-order validation as comparison of model predictions *with existing data from studies that were not used to parameterize, construct, or develop the model*. This could include, for example, comparison of model predictions with results of randomized trials that were not used to develop model inputs. Finally, we defined predictive validity as comparison of model predictions to previously unobserved data collected subsequent to and independent of model development, such as comparison of model results to clinical trials conducted after the model's inception. Predictive validation is so named because it involves comparison of the model's predicted outcomes with future patterns of health outcomes—that is, it is intended to gauge the ability of the model to *predict* future outcomes of interest. Obviously, predictive validation is only possible over time as data sources (such as clinical trials or reports of national trends in disease incidence) become available to which model results can be compared.

In characterizing model validation methods, we describe attributes of validation techniques including type of validation, health outcome endpoints used as the basis for comparing modeled to observed data, and number and type of studies from which data for comparisons are obtained, referred to as *comparator studies*. We describe methods by which comparisons were made between modeled and observed data (e.g., simulation of a trial cohort using the model) and criteria by which adequacy of validation is judged, noting the concordance metrics and statistical tests used, if any.

RESULTS

We identified 13 disease simulation models, six of asthma (12–17) and seven of COPD (18–24). Selected characteristics of the models are provided in Table 1. A majority (62%) were from Europe and funded by pharmaceutical companies (70%). Eleven were state-transition models; two were dynamic population models. Nine models were developed for specific cost-utility analyses (CUA) and four for exploratory analyses or future CUA. A summary of the types of validation used with these models is given in Table 2.

Eleven models included no mention of first-order validation; one did so implicitly and one explicitly. Seven (54%) models included tests of second-order validity, in which model results are evaluated for their agreement with data from studies used in the development of the model (Table 3). Most of the seven studies reporting tests of second-order validity appeared to use

simulations of source study cohorts that compared modeled with observed outcomes (referred to hereafter as *cohort simulation*), although only two of the five explicitly described the validation method in this manner. A variety of health outcome endpoints was used for tests of second-order validity, including disease exacerbation, time spent in specific disease states, number of patients in specific disease states, COPD prevalence, and all-cause mortality. Most second-order validation results were expressed as qualitative comparisons of a point estimate generated by the model with a point estimate of the analogous endpoint from the comparator study. Only one reported a statistical test of modeled versus observed endpoints (12); otherwise, *a priori* criteria for judging the acceptability of validation were not reported.

An illustrative second-order validation is that reported by Oostenbrink et al. (22), in which clinical trial data were used to parameterize the model (e.g., to generate transition probabilities between disease states), after which the model was populated with a hypothetical cohort of patients with baseline characteristics matching the trial patients. The model was run with this simulated cohort and a time horizon approximating the trial, and model-predicted outcomes were compared to outcomes observed in the trial. Oostenbrink et al. compared, for example, the model-predicted and observed mean number of COPD exacerbations over a 6-month time horizon, finding that the model predictions “were somewhat higher” than the mean number of exacerbations observed in the trial.

Seven of thirteen (54%) models included tests of third-order validity, in which modeled outcomes are assessed for agreement with results of studies not used to develop the model (Table 4). These models overlapped with but were not identical to the group of seven models reporting tests of second-order validity. A variety of validation endpoints was used including COPD prevalence, exacerbation rate, severity distribution, and mortality. Six of seven models reporting third-order validation did so by qualitatively contrasting modeled outcomes with published data. One used a cohort simulation to quantitatively assess agreement between model outcomes and the comparator study. Validation was typically described as acceptable in qualitative terms despite absence of criteria for judging adequacy of validation.

An illustrative third-order validation is that reported by Borg et al. (18), in which the model-predicted annual COPD exacerbation rate was compared qualitatively to that reported in a randomized trial and to a retrospective study of COPD patients. Although Borg et al. do not report the exacerbation rates from the comparator studies, they conclude that the model-predicted “frequency of exacerbations... becomes 1.27 per patient per year, which is in line with published data.” No models included tests of predictive validity or plans for predictive validation.

DISCUSSION

Of the 13 obstructive lung disease simulation models we identified in the recent literature, ten had attempted model validation by assessing model outcomes for agreement with data from published studies (using what we refer to as second- or

Table 1. Selected characteristics of obstructive lung disease simulation models

Reference	Country	Purpose	Funder	Method	Health outcomes	No. states	Cycle length	Time horizon
					<i>Asthma simulation models</i>			
Combescurre et al., 2003	France	Exploratory ^a	Pharm	state-transition	Duration of asthma control Rates of transition from control to uncontrolled asthma	3	1 d	400 d
Fuhlbrigge et al., 2006 ^b	US	CUA	Govt Pharm ^b	state-transition	QALM Symptom days Asthma mortality Asthma exacerbations Rate of hip and other fracture Hip and other fracture mortality	NR	1 mo	10 yr
Paltiel et al., 2001	US	CUA	Pharm	state-transition	QALM Symptom-free days All-cause mortality Asthma mortality Asthma exacerbations	123 ^c	1 mo	10 yr
Price et al., 2002	UK	CUA	Pharm	state-transition	Controlled weeks QALY	5	1 wk	12 wk
Saint-Pierre et al., 2003	France	Exploratory ^a	Pharm Govt Prof Org	state-transition	Asthma exacerbations	3	1 d	400 d
Wild et al., 2005	US	CUA	Govt	state-transition	QALY Symptom days Symptom-free days Disability cases	NR	1 mo	10 yr
					<i>COPD simulation models</i>			
Borg et al., 2004	Sweden	Future CUA ^d	Pharm	state-transition	LY QALY Time without exacerbation COPD exacerbations	28 ^e	1 yr 1 wk	Lifetime
Feenstra et al., 2001	Netherlands	Exploratory ^f	Govt	dynamic population model	LY DALY COPD prevalence	n/a	1 yr	1994 — 2015

(Continued on next page)

Hoogendoorn ^g et al., 2005	Netherlands	Exploratory CUA	Govt	dynamic population model	LY QALY COPD prevalence by severity Mortality	n/a	1 yr	2000 — 2025
Johansson et al., 2005	Sweden	CUA	NR	state-transition	COPD rate per 10,000 COPD mortality All-cause mortality Years of Life Saved (YLS) QALY	8		Lifetime
Oostenbrink et al., 2005	Netherlands	CUA	Pharm	state-transition	QALM COPD exacerbations	g ^h	8 d 1 mo	1 yr
Sin et al., 2004	Canada	CUA	Pharm Govt	state-transition	QALY All-cause mortality COPD exacerbations	3	3 mo	3 yr
Spencer et al., 2005	Canada	CUA	Pharm	state-transition	Minor COPD exacerbations Major COPD exacerbations All-cause mortality adjusted for COPD severity QALY	4	3 mo	25 yr

CUA = cost-utility analysis; d = day; DALY = disability-adjusted life years; Govt = government; LY = life years; mo = month; NR = not reported; Pharm = pharmaceutical company; Prof Org = professional organization or society; QALM = quality-adjusted life months; QALY = quality-adjust life years; wk = week; yr = year.

^aThe purpose of the models reported by Combes et al. and Saint-Pierre et al. is to investigate long-term asthma control.

^bThe model reported by Fuhlbrigge et al. is an extension of the Asthma Policy Model reported by Paltiel et al., 2001.

^cAlthough the number of states is unclear from the manuscript describing the model, the description of the transition matrix in the manuscript appendix implies 123 states.

^dThe purpose of the model reported by Borg et al. is to document a disease simulation model for COPD useful for economic evaluations of new medicines.^e

^eIncludes seven primary states with four states embedded in each, for assumed total of 28 states.

^fThe purpose of the model reported by Feenstra et al. is to generate exploratory projections of future COPD burden.

^gThe model reported by Hoogendoorn et al. is an extension of the model reported by Feenstra et al., 2001.

^hIncludes three primary states with three states embedded in each, for assumed total of nine states.

third-order validation). Comparing modeled outcomes with data from studies used in model development was the method of validation for seven of the models. In its simplest and most common form, this method involved comparing modeled outcomes to observed outcomes from the same data set used to develop the model (e.g., by matching or recreating the trial cohort). Another form of second-order validation involved comparing modeled outcomes to observed outcomes from a subset of data from studies used in model development but intentionally removed from model parameterization and “set aside” for validation. This splitting of a single population sample to facilitate validation seems an unlikely method to identify underlying biases in the model and carries the disadvantageous of reducing study power for model parameterization. Comparing modeled outcomes with data from studies used in model development (including the split sample method) represents a lower standard of validation than comparison to data independent of the model.

For both types of validation methods (i.e., those using data from studies used in model development and those using data independent of model development), a wide variety of validation endpoints was used, with little justification given for endpoint selection. It appeared that, for most models, the choice of validation endpoint was dictated by availability of endpoints in the comparator study, but this explanation was seldom explicit. Similarly, various types of comparator studies provided data to assess model predictions, but justification for choice of comparator studies was generally weak or absent. For several models, the exclusion of data that otherwise could have been used for validation was unexplained. There was near-universal absence of criteria for assessing adequacy of validation.

Currently there exists no clear consensus as to what constitutes an appropriate level of model validation for disease simulation models. However, recent guidelines suggest an emerging consensus that models should, at a minimum, demonstrate what we have called second-order validity (that is, agreement with data from studies used in model development), with failure to do so strongly suggestive “that the structure of the model is faulty” (6, 11) and should strive for third-order validity (agreement with data from studies not used in model development) (6–8). There is disagreement about the need for models to demonstrate predictive validity, with some experts contending that using models based on historical data to predict future outcomes is meaningless, since such models cannot be expected to reflect unknown future conditions, while others believe that predictive validity is valuable but not essential. Certainly there are undeniable practical impediments to undertaking tests of predictive validity given that years of data would need to be accumulated before model predictions could be compared to observed outcomes.

If we assume second-order validity as the minimum level of appropriate validation, about half of the obstructive lung disease simulations models we identified demonstrated appropriate validation. However, even those that reported second-order validation tended to fail to establish clear validation goals and criteria for judging adequacy of validation. If we include all types of reported second- or third-order validation, regardless of the methodology or quality of the comparison involved, about three-quarters of the models could be said to include some sort of validation. At first glance, this proportion appears favorable, suggesting that most obstructive disease simulation models are indeed validated. However, this proportion should be interpreted with caution given that we used a broad definition of validation

Table 2. Summary of types of validation used in obstructive lung disease simulation models

Reference	1st order validation (debugging)	2nd order validation (compared to existing data used in model development)	3rd order validation (compared to existing data not used in model development)	Predictive validation (compared to future data independent of model)
<i>Asthma simulation models</i>				
Combesure et al., 2003	Not described	YES	NO	NO
Fuhlbrigge et al., 2006	Not described	NO	YES	NO
Paltiel et al., 2001	Implicit	NO	NO	NO
Price et al., 2002	Not described	NO	NO	NO
Saint-Pierre et al., 2003	Not described	YES	NO	NO
Wild et al., 2005	Not described	NO	YES	NO
<i>COPD simulation models</i>				
Borg et al., 2004	Not described	YES	YES	NO
Feenstra et al., 2001	Not described	YES	YES	NO
Hoogendoorn et al., 2005	Not described	YES	YES	NO
Johansson et al., 2005	YES	NO	YES	NO
Oostenbrink et al., 2005	Not described	YES	NO	NO
Sin et al., 2004	Not described	NO	NO	NO
Spencer et al., 2005	Not described	YES	YES	NO

YES = this type of validation was used for at least one outcome but not necessarily for all outcomes.

NO = this type of validation was not used for any outcome.

Table 3. Second-order validation methods used in recent obstructive lung disease simulation models

Reference	Validation endpoint(s)	Comparator study	Method	Concordance metric	Assessment of validity
<i>Asthma simulation models</i>					
Combescuré et al., 2003	Number of patients in mild, moderate, severe asthma states at 100, 200, 300, and 400 days	3-yr prospective study of 365 French adult asthmatics	Cohort simulation	Chi-square statistic of observed versus modeled number of patients in each state (actual modeled number of patients at each time point not reported)	The number of patients in each control state was not significantly different from the theoretical distribution expected from the model... ¹⁹ This study demonstrates that a Markov model can be developed and used to predict the long-term control of asthma. ²⁰ The results (not shown) seemed to indicate that the models could be seen as tied to the data. ²¹
Saint-Pierre et al., 2003*	Number of patients in optimal, suboptimal, and unacceptable asthma control states	4-yr prospective study of 459 French adult asthmatics	Not explicit—appears to be cohort simulation	Qualitative comparison of observed to modeled number of patients in each state (actual counts not reported)	
<i>COPD simulation models</i>					
Borg et al., 2004	(a) All-cause mortality (b) Time spent in GOLD Stages I, IIA, IIB, III (c) Mortality among severe patients during severe exacerbation (d) Mortality among patients with FEV ₁ <80% predicted during severe exacerbation	(a) and (b) 10-year follow-up study of 1457 Swedish COPD patients (c) Prospective cohort of 1016 COPD patients (d) Prospective study of 83 COPD patients	Not explicit—appears to be cohort simulation	(a) Qualitative comparison of modeled probability of death (0.12) to observed (0.55) (b) Qualitative comparison of time spent in GOLD stages (not quantified) (c) Qualitative comparison of modeled (11.9%) versus reported (11%) mortality during severe exacerbation (d) Qualitative comparison of modeled (4.87%) versus reported (4%) mortality during severe exacerbation	There remains mainly two large discrepancies: mortality in Stage I, which is overestimated by the model, and the probability of remaining in Stage IIA over 10 years, which is underestimated by the model... Apart from these two large discrepancies and discrepancies owing to the scarce number of observations or modeling assumptions, the model reproduces the observed outcomes fairly well. ²²
Feenstra et al., 2001	COPD prevalence per 1000, 1995–1998	Continuous morbidity registration from general practitioners, 1995–1998	Not reported	Qualitative comparison of modeled versus observed annual COPD prevalence and difference in modeled versus observed COPD prevalence for men (0.66) and women (3.52)	The model predicts somewhat lower prevalence rates, but the differences in prevalence rates between the model and the CMR [comparator study]... should be seen as an indication that there are no large biases in the model for short-term projections. ²³

(Continued on next page.)

Hoogendoorn et al., 2005	COPD prevalence per 1000, 2000–2003	Continuous morbidity registration from general practitioners, 2000–2003	Not reported	Qualitative comparison of modeled versus observed difference in annual COPD prevalence for men (3.71) and women (0.42)	...it was concluded that the current model projections compare quite well with this GP registration."
Oosterbrink et al., 2005	Mean (SE) of COPD exacerbations	Multi-center randomized trial of tiotropium in COPD patients	Cohort simulation	Qualitative comparison of modeled versus observed mean (SE) and difference between treatment groups	A comparison... showed that the mean estimates of the model were somewhat higher than observed in the trials, but that the estimated difference between treatment groups was exactly the same." It was shown that the model closely resembled the difference in the numbers of exacerbations what were observed in the original trials. The ability to compare the outcomes of the model with the original trial data makes the model transparent and may thereby increase the acceptance of this model..."
Spencer et al., 2005	COPD exacerbation rate	Multi-center randomized trial of combined salmeterol and uticasone in COPD patients	Not explicit —appears to be cohort simulation	Qualitative comparison of modeled (1.3) versus observed COPD exacerbations per year (observed COPD exacerbations in comparator study not reported)	The model predicts a background rate of around 1.3 exacerbations per year, thus replicating exactly the rate seen in the TRISTAN study from which the main clinical results are taken and to which the model is applied."

*The model reported by Saint-Pierre et al. is an enhancement (to include time-varying transition probabilities) of the model reported by Combescurre et al. COPD = Chronic Obstructive Pulmonary Disease; GOLD = Global Initiative for Chronic Obstructive Lung Disease; SE = standard error of the mean.

Table 4. Third-order validation methods used in recent obstructive lung disease simulation models

Reference	Validation endpoint(s)	Comparator study	Method	Concordance metric	Overall assessment of validity
<i>Asthma simulation models</i>					
Fuhrigge et al., 2006	Odds ratio of fracture in ICS-treated versus untreated patients	1 case-control study	Cohort simulation	Qualitative comparison of odds ratio of fracture	... estimates were similar to those reported by Hubbard et al."
Wild et al., 2005	Prevalence of isocyanate induced asthma	Four studies, primarily occupational cohort studies	Qualitative comparison of modeled and published estimates	Qualitative comparison of modeled occupational isocyanate asthma prevalence (25%) to published occupational isocyanate asthma prevalence (as high as 30%)	NR
<i>COPD simulation models</i>					
Borg et al., 2004	(a) All-cause mortality (b) Annual COPD exacerbation rate	(a) National mortality rate (b) Randomized controlled trial of 173 COPD patients and retrospective cohort study of 107 COPD patients	Qualitative comparison of modeled and published estimates	(a) Qualitative comparison of mortality rates (b) Qualitative comparison of modeled versus reported annual COPD exacerbation rate	... frequency of exacerbations... then becomes 1.27 per patient per year, which is in line with published data."
Feenstra et al., 2001	COPD prevalence in 1997	Medication data from Dutch sickness fund" in 1997	Qualitative comparison of modeled and published estimates	Qualitative comparison of modeled 1997 prevalence (262,000 cases) to observed 1997 prevalence (250,000 to 300,000 cases)	This compares well with an estimate based on medication data from Dutch sickness funds..."
Hoogendoorn et al., 2005	COPD severity distribution	1 cross-sectional population-based study of 317 patients with COPD from Maastricht	Qualitative comparison of modeled and published estimates	Qualitative comparison of modeled COPD severity distribution in 2003 (29% mild, 52% moderate, 16% severe, 3% very severe) to observed 2002/2003 severity distribution (30% mild, 48% moderate, 17% severe, 5% very severe)	... they were quite close to the estimates from Maastricht"
Johansson et al., 2006	(a) Smoking-related mortality (b) Difference in all-cause mortality between smokers and quitters (c) COPD mortality (d) YLS (e) QALYs	(a) Large prospective Danish study and large prospective British study (b) U.S. study of all-cause mortality (c) Disease simulation model of smoking-related mortality (d) economic evaluation review paper by UK National Institute for Health and Clinical Excellence (e) decision analytic model of nicotine patch for smoking cessation	Qualitative comparison of modeled and published estimates	(a) Qualitative comparison of modeled smoking-related mortality (55-60%) to published smoking-related mortality (50%) (b) Qualitative comparison of modeled decreased risk of death (10-15%) to published values (not reported) (c) Qualitative comparison of COPD mortality, noting only that model results are contrary to results of comparator study (d) Qualitative comparison of modeled YLS (0.00 to 2.24, mean 1.09) with published values (0.28-2.4) (e) Qualitative comparison of modeled QALYs gained (-0.01 to 0.55) with published values (1.08 to 1.94)	Decreased risk of death... is in the range of a US study on all-cause mortality... " The model estimates of YLS thus seems to be consistent with studies that only include highly smoking-related diseases." Regarding QALY endpoint, This difference, apart from the underlying difference in life-years saved, is explained by the QoL-weights chosen." Overall, The model simulations however seem to be in accordance with previous studies on the health effects of tobacco cessation, when only the most smoking-related diseases are included."

(Continued on next page.)

Spencer et al., 2005	COPD exacerbation rate	1 1-year cohort of 70 COPD patients	Qualitative comparison of modeled and published estimates	Qualitative comparison of modeled COPD exacerbation rate (1.3 per year) to published values (not reported)	Observational data suggest that this rate may be lower than that seen in the community. ^a While we believe that the assumptions underlying the model's predictions of life expectancy are robust and the resulting predictions appear reasonable, we are not aware of any direct observations of a similar population over a sufficient period to test these predictions. ^a
----------------------	------------------------	-------------------------------------	---	--	---

COPD = Chronic Obstructive Pulmonary Disease; ICS = inhaled corticosteroid; QALY = Quality-Adjusted Life Year; YLS = Years of Life Saved; QoL = Quality of Life; NR = Not reported.

that included any attempt by model developers to compare modeled outcomes with other data, regardless of whether that activity was labeled *validation*. Although we believe it an encouraging finding that over half of the models report some sort of validation, this result does not imply that current validation methods are sufficiently developed, implemented, or documented.

As recognized in recent guidance on use of disease simulation models (7), some notion of validity is needed to protect against incorrect conclusions and to protect users from being misled by disease simulation models. To promote acceptance of disease simulation models by decision-makers in the clinical and public health communities, there need to be intuitively reasonable and consistently reported validation goals. Our review of validation methods used with recent obstructive lung disease models suggests that the following activities may help achieve those goals. Specifically, experts in model development and end-users of disease simulation models in the clinical and public health communities could collaborate to

1. Establish standard terminology for describing validation methods, including explicit definition of what constitutes *calibration* versus *validation*, and *internal* versus *external* and *dependent* versus *independent* data.
2. Define levels of validation recommended in various decision contexts, which involves addressing how various uses of obstructive lung disease (or other disease) simulation models influence the stringency of validation needed to improve model credibility and acceptance.
3. Explore in detail the rationale for and conflicting opinions on the need for predictive validity. Provide guidance on the circumstances in which the value of the information provided by predictive validation justifies the time and expense needed to undertake it.
4. Develop *practical* guidelines for best practices in the conduct of validation. For example, practical guidelines for selection of validation endpoints and *a priori* definition of criteria for judging adequacy of validity (including concordance metrics and statistical testing methods) would encourage model developers to conduct and report validation in a systematic manner.
5. Encourage standardized and detailed reporting of model development. This is similar to the recommendation by the Panel on Cost-Effectiveness in Health and Medicine (25) that cost-effectiveness analyses be comprehensively described in a "technical report" providing analytic detail beyond what can typically be included in most journal reports. A similar technical report detailing model development, including validation, would increase availability of information needed to understand and evaluate disease simulation models.
6. Standardize elements of validation reporting. Guidelines could specify that reports of validation include, at a minimum, descriptions of the type of validation, comparator studies and justification for their selection, methodologic details of how the comparison was performed (e.g., by describing cohort characteristics included in the model

simulation) and quantitative side-by-side comparison of actual model-predicted and observed endpoints.

7. Establish a "data resource library" of publicly-available datasets useful for validation of obstructive lung disease and other disease simulation models. These datasets would constitute a "gold standard" for comparison, and journals reporting analyses based on obstructive lung disease simulation models could be encouraged to mandate validation using them.

Finally, we acknowledge that a thorough assessment of model quality and credibility would include evaluation of a broader array of model characteristics, including model structure and quality of model inputs. An evaluation of these characteristics, while important, is outside the scope of this review.

In conclusion, we found that over half of recent obstructive lung disease simulation models report validation of some sort. However, the lack of detailed reporting of validation results and inconsistencies in validation methodology and interpretation make the assessment of validation difficult. For simulation modeling to be accepted as a tool for evaluating clinical and public health programs, models must be validated to credibly simulate health outcomes of interest. Defining the required level of validation and providing guidance for quantitative assessment and reporting of validation are important future steps in promoting simulation models as practical decision tools.

REFERENCES

1. Mannino DM, Homa DM, Akinbami LJ, Ford ES, Redd SC. Chronic obstructive pulmonary disease surveillance United States 197 —2000. *Morbidity and Mortality Weekly Report* 2002; 51:1—16.
2. Mannino DM, Homa DM, Akinbami LJ, Moorman JE, Gwynn C, Redd SC. Surveillance for Asthma United States, 1980—1999. *Morbidity and Mortality Weekly Report* 2002; 51:1—13.
3. U.S. Centers for Disease Control and Prevention. Asthma: General Information. 4-6-2007. Atlanta, US Centers for Disease Control and Prevention. Report.
4. U.S. Department of Health and Human Services. Facts about Chronic Obstructive Pulmonary Disease (COPD). 4-6-2007. Atlanta, US Centers for Disease Control and Prevention. Report.
5. U.S. Department of Health and Human Services. Healthy People 2010: Understanding and Improving Health, 2nd Ed. 017-001-001-00-550-9. 11-1-2000. Washington DC, US Government Printing Office Report.
6. Philips Z, Ginnelly L, Sculpher M, Claxton K, Golder S, Rieksma R, et al. Review of guidelines for good practice in decision-analytic modelling in health technology assessment. *Health Technol Assess* 2004; 8:iii—xi,1.
7. Weinstein MC, Toy EL, Sandberg EA, Neumann PJ, Evans JS, Kuntz KM, et al. Modeling for health care and other policy decisions: uses, roles, and validity. *Value Health*. 2001; 4:348—361.
8. Weinstein MC, O'Brien B, Hornberger J, Jackson J, Johannesson M, McCabe C, 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.
9. Buist AS, Vollmer WM, Sullivan SD, Weiss KB, Lee TA, Menezes AM et al. The Burden of Obstructive Lung Disease Initiative (BOLD): rationale and design. *COPD J* 2005; 2:277—283.

10. Lee TA, Rutten-van Molken MP. Economic modeling in chronic obstructive pulmonary disease. *Proceedings of the American Thoracic Society* 2006; 3:630—634.
11. Eddy DM. Technology Assessment: The Role of Mathematical Modeling. In: Mosteller F, ed. *Assessing Medical Technologies*. Washington, DC: National Academy Press; 1985:144—160.
12. Combescore C, Chanez P, Saint-Pierre P, Daures JP, Proudhon H, Godard P. Assessment of variations in control of asthma over time. *Eur Respir J* 2003; 22:298—304.
13. Fuhlbrigge AL, Bae SJ, Weiss ST, Kuntz KM, Paltiel AD. Cost-effectiveness of inhaled steroids in asthma: impact of effect on bone mineral density. *J Allergy Clin Immunol* 2006; 117:359—366.
14. Paltiel AD, Fuhlbrigge AL, Kitch BT, Liljas B, Weiss ST, Neumann PJ, et al. Cost-effectiveness of inhaled corticosteroids in adults with mild-to-moderate asthma: results from the asthma policy model. *J Allergy Clin Immunol* 2001; 108:39—46.
15. Price MJ, Briggs AH. Development of an economic model to assess the cost effectiveness of asthma management strategies. *Pharmacoeconomics* 2002; 20:183—194.
16. Saint-Pierre P, Combescore C, Daures JP, Godard P. The analysis of asthma control under a Markov assumption with use of covariates. *Stat Med* 2003; 22:3755—3770.
17. Wild DM, Redlich CA, Paltiel AD. Surveillance for isocyanate asthma: a model based cost effectiveness analysis. *Occup Environ Med* 2005; 62:743—749.
18. Borg S, Ericsson A, Wedzicha J, Gulsvik A, Lundback B, Donaldson GC et al. A computer simulation model of the natural history and economic impact of chronic obstructive pulmonary disease. *Value Health* 2004; 7:153-67.
19. Feenstra TL, van Genugten ML, Hoogenveen RT, Wouters EF, Rutten-van Molken MP. The impact of aging and smoking on the future burden of chronic obstructive pulmonary disease: a model analysis in the Netherlands. *Am J Respir Crit Care Med* 2001; 164:590—596.
20. Hoogendoorn M, Rutten-van Molken MP, Hoogenveen RT, van Genugten ML, Buist AS, Wouters EF, et al. A dynamic population model of disease progression in COPD. *Eur Respir J* 2005; 26:223—233.
21. Johansson PM, Tillgren PE, Gulbrandsson KA, Lindholm LA. A model for cost-effectiveness analyses of smoking cessation interventions applied to a Quit-and-Win contest for mothers of small children. *Scand J Public Health* 2005; 33:343—352.
22. Oostenbrink JB, Rutten-van Molken MP, Monz BU, FitzGerald JM. Probabilistic Markov model to assess the cost-effectiveness of bronchodilator therapy in COPD patients in different countries. *Value Health* 2005; 8:32—46.
23. Sin DD, Golmohammadi K, Jacobs P. Cost-effectiveness of inhaled corticosteroids for chronic obstructive pulmonary disease according to disease severity. *Am J Med* 2004; 116:325—331.
24. 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—637.
25. Gold MR, JE Siegel, LB Russell, MC Weinstein. *Cost-Effectiveness in Health and Medicine*. New York: Oxford University Press; 1996.

SUPPLEMENTAL MATERIAL

Search terms and number of abstracts and eligible papers obtained from each search are given here.

Primary Search

Search #1: (*asthma* OR *COPD*) AND (*cost-effectiveness* OR *Markov* OR *decision analysis*) yielded 482 abstracts, of which 23 met screening criteria and 10 met inclusion criteria.

Secondary Searches

Search #2: (*asthma* OR *COPD*) AND (*natural history*) AND (*Markov* OR *decision analysis*) yielded 4 abstracts, of which 3 met screening criteria and 3 met inclusion criteria. All three abstracts had been previously identified by Search #1.

Search #3 (*asthma* OR *COPD*) AND (*life table model*) yielded 53 abstracts, of which 12 met screening criteria, nine met inclusion criteria, and six were previously identified (by Search #1), hence yielding three additional papers.

Search #4: (*asthma* OR *COPD*) AND (*model*) AND (*validation*) yielded 41 abstracts, of which one met screening criteria, one met inclusion criteria and had been previously identified by Search #1.

Total

Ten publications from Search #1 and three publications from Search #3 for a total of 13 publications are included in our review.