Volume 10 • Number 5 • 2007
VALUE IN HEALTH

Economic Evaluation of a Bayesian Model to Predict Late-Phase Success of New Chemical Entities

Asher D. Schachter, MD, MMSc, MS,[1,2,3] Marco F. Ramoni, PhD,[2,3,4] Gianluca Baio, PhD,[5,6] Thomas G. Roberts, Jr, MD, MSocSci,[7] Stanley N. Finkelstein, MD[8]

[1]Children's Hospital Boston, Boston, MA, USA; [2]Children's Hospital Informatics Program at Harvard—MIT Health Sciences and Technology, Boston, MA, USA; [3]Harvard Medical School, Boston, MA, USA; [4]Harvard Partners Center for Genetics and Genomics, Boston, MA, USA; [5]University College London, London, UK; [6]Fondazione CERM, Rome, Italy; [7]Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA; [8]Massachusetts Institute of Technology, Cambridge, MA, USA

ABSTRACT

Objective: To evaluate the economic impact of a Bayesian network model designed to predict clinical success of a new chemical entity (NCE) based on pre-phase III data.

Methods: We trained our Bayesian network model on publicly accessible data on 503 NCEs, stratified by therapeutic class. We evaluated the sensitivity, specificity and accuracy of our model on an independent data set of 18 NCE-indication pairs, using prior probability data for the antineoplastic NCEs within the training set. We performed Monte Carlo simulations to evaluate the economic performance of our model relative to reported pharmaceutical industry performance, taking into account reported capitalized phase costs, cumulative revenues for a postapproval period of 7 years, and the range of possible false negative and true negative rates for terminated NCEs within the pharmaceutical industry.

Results: Our model predicted outcomes on the independent validation set of oncology agents with 78% accuracy (80% sensitivity and 76% specificity). In comparison with the pharmaceutical industry's reported success rates, on average our model significantly reduced capitalized expenditures from \$727 million/successful NCE to \$444 million/successful NCE ($P < 0.001$), and significantly improved revenues from \$347 million/phase III trial to \$507 million/phase III trial ($P < 0.001$) during the first 7 years post launch. These results indicate that our model identified successful NCEs more efficiently than currently reported pharmaceutical industry performances.

Conclusions: Accurate prediction of NCE outcomes is computationally feasible, significantly increasing the proportion of successful NCEs, and likely eliminating ineffective and unsafe NCEs.

Keywords: Bayesian network, Monte Carlo simulation, new chemical entity, pharmacoeconomics, prediction.

Introduction

The estimated cost of developing a single new chemical entity (NCE) into a successful therapeutic agent is US\$802 million (in year 2000 dollars, 95% confidence interval of \$684–936 million) [1], with clinical phase costs of \$467 million, and taking into account "time costs" related to the length of time from Investigational New Drug (IND) approval to New Drug Application (NDA) marketing approval, and the cost of capital. Although the true cost of NCE development has generated debate [2–6], innovative drug development is expensive by any measure. A substantial component of this expense is related to the proportion of NCEs that fail during the clinical trial development phase [7,8], particularly those NCEs that fail during or after the more costly later phases.

Over 60% of terminations occur later in the drug development process, during phases II and III [9]. Because phase III studies generally require larger and therefore more costly clinical trials, earlier termination of even a fraction of late-phase failures gives rise to a factoring of savings: terminating only 5% of all phase III clinical failures in phase I would reduce out-of-pocket clinical costs by 5.5–7.1% [10]. It is therefore in the pharmaceutical industry's interest to terminate failures early, without compromising the quality of the clinical trials or terminating successful agents, and most importantly, without exposing clinical trial subjects to unnecessary risk.

Address correspondence to: Asher Schachter, Division of Nephrology, Children's Hospital Boston, 300 Longwood Avenue, Boston, MA 02115, USA. E-mail: asher.schachter@childrens.harvard.edu

10.1111/j.1524-4733.2007.00191.x

Financial interest disclosures: Dr. Marco Ramoni is a member of the Board of Directors of Bayesware, LLC. Dr. Thomas Roberts is currently an employee of Noonday Asset Management, has no conflicts of interest to disclose. Dr. S. Finkelstein is the principal investigator on a grant from Merck, Co., Inc., is a consultant to Thomson Medstat. The Massachusetts Institute of Technology Program on the Pharmaceutical Industry receives support from the Sloan Foundation and from the Merck Foundation.

We have constructed a Bayesian network model [11,12] to calculate the probability that a specific NCE will be clinically successful through phase III/NDA approval given 1) prior data regarding success rates for NCEs of the same therapeutic class (e.g., antineoplastic, cardiovascular) and source (i.e., licensed in vs. developed in-house, USA vs. non-USA); and 2) the NCE's pre-phase III toxicology and efficacy data. The aim of our model is to predict clinical success (safety and efficacy) for a specific NCE in question, given prior data on NCEs in the same therapeutic class. To evaluate our model's performance, we compiled an independent data set of oncology NCEs from several public sources, and performed a Monte Carlo simulation, sampling costs and revenues from reported pharmacoeconomic data distributions, to determine the economic impact of our model relative to reported pharmaceutical industry performances.

Methods

Sources of Data

Training data set. The data used to populate the conditional probability tables for the Bayesian model [11,12] were obtained from the Tufts Center for the Study of Drug Development (TCSDD) sources [9]. The TCSDD data included a total of 503 NCEs divided more than 10 therapeutic classes, as defined by the TCSDD investigators, including Analgesic/Anesthetic (49 NCEs), Antimicrobial (57 NCEs), Antineoplastic (38 NCEs), Cardiovascular (120 NCEs), Central Nervous System (110 NCEs), Endocrine (33 NCEs), Gastrointestinal (15 NCEs), Immunologic (13 NCEs), Respiratory (25 NCEs), and Miscellaneous (43 NCEs). These data were descriptive only and did not include information on individual NCEs.

Data on terminated NCEs. DiMasi [9] analyzed the causes of failure for 348 NCEs that were withdrawn from development. Nevertheless, NCEs that proceeded through all clinical trial phases but failed to achieve NDA approval are not included in DiMasi's analysis, and data on terminated NCEs were not stratified by therapeutic class. Furthermore, DiMasi stratified by "primary" cause of failure, thereby not disclosing any degree of overlap, that is, NCEs that failed primarily for one reason, but may have also failed for another reason (e.g., an NCE that failed because it was not safe, but was also not effective). The DiMasi analysis demonstrated that of a total of 348 NCEs that were terminated, the primary reason for termination was efficacy in 121, safety in 72, economics in 109, and "other" in 46. Assuming that the proportions of causes of failure are consistent across the withdrawn drugs, we used these values to calculate an "overlap function" that accounts for the probability that a given NCE will fail because of both safety and efficacy concerns [11,12].

Independent data set collection. We approached investigators at five pharmaceutical companies, requesting from each de-identified data, including data on terminated NCEs, to evaluate our model, but we were unsuccessful in obtaining relevant data. Therefore, we constructed an independent data set starting from a collection of 213 published clinical trials previously compiled from American Society of Clinical Oncology (ASCO) abstracts of phase I NCEs [13]. We then acquired outcome data (latest phase of development, ongoing clinical trial activity, NDA approval, phase at termination if relevant) for all antineoplastic NCEs contained within the PharmaProjects database. We determined which NCEs were common to both the ASCO-derived database and the PharmaProjects database. To acquire data regarding the characteristics of these NCEs, we searched the National Center for Biotechnology Information PubMed database for toxicity and early-phase efficacy data for each NCE-indication pair (e.g., an NCE directed against a specific cancer type). We were able to extract data for 18 NCE-indication pairs, for which half were consistent across data inputs in that all efficacy studies for the given NCE were performed on the same cancer type. The remaining half of NCE-indication pairs was mixed, with different cancer types studied across the in vitro, in vivo, and phase I/II studies. For each NCE-indication pair, the independent data set lists the lowest reported therapeutic index for a vital organ, the lowest reported therapeutic index for a disease-related complication, and reported data on in vitro, in vivo, and pre-phase III human efficacy studies. When multiple studies were available, the earliest reported data were used. We performed an additional search of public sources to determine whether any NCEs listed as "active" in our database have recently been approved. This search resulted in changing one NCE that was listed as "active phase III" in our database to "marketed." Half of the NCE-indication pairs achieved either NDA approval or were still under active investigation in phase III, whereas the remaining half of NCE-indication pairs were either terminated at or before phase III, or were active at a phase before phase III.

Bayesian network model. A Bayesian Network (BN) is a directed acyclic graph consisting of nodes (representing stochastic variables) and arrowed arcs (representing dependencies between the variables) [14,15]. A BN has the capability of computing the probability that the event represented by the top (root) node will occur, given the data encoded within the lower (hidden) and bottom (leaf) nodes. Data are entered for the leaf

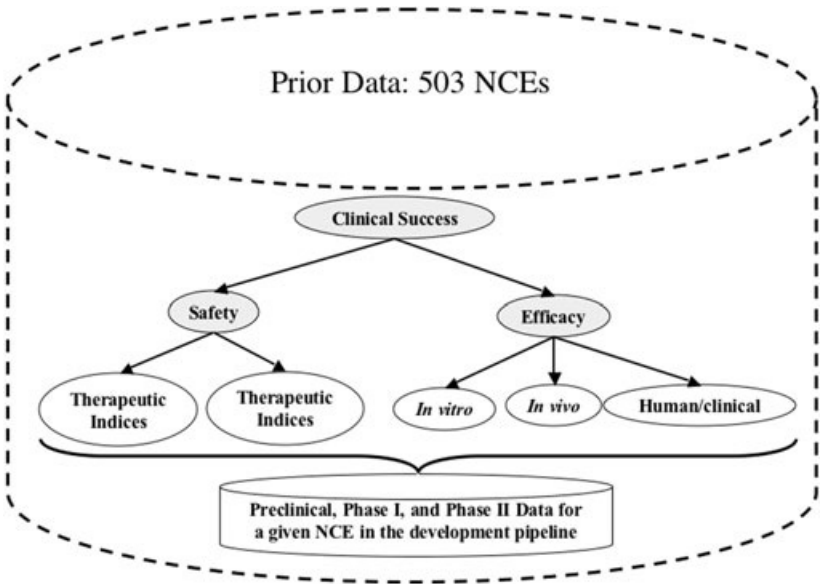

Figure 1 Bayesian network (BN) engineered specifically to predict clinical success for a new chemical entity (NCE). Shaded nodes represent stochastic variables, and unshaded nodes represent input variables, which are fed data from preclinical, phase I, and phase II studies for an NCE under development. The BN computes the probability of clinical success for the NCE in question given prior data on 503 NCEs, stratified by therapeutic class.

nodes, and Bayes theorem is used to "climb" up the tree, to calculate the posterior probability of the root node's state given the data. We engineered a BN model [11,12] encompassing the key variables in clinical drug development (Fig. 1). Briefly, our BN model [11,12] computes the probability of clinical success for an NCE, given the NCE's characteristics (therapeutic class, development source, i.e., in-licensed vs. in-house, therapeutic indexes, and in vitro, in vivo, and clinical data), in the context of prior data [1] on NCEs in the same therapeutic class.

Implementation

We implemented the model in the Java 2 programming language, within Apple Xcode on Apple OS X Tiger, version 10.4.2 (Apple Inc., Cupertino, CA). An object-oriented, model-view approach [16] was used to structure the program.

Model Evaluation

We had two prior probability models, one using a pessimistic prior probability and one using an optimistic prior probability. The rationale for using pessimistic and optimistic prior probabilities is that the TCSDD reports (from which we derived our conditional probability tables) provide the probability of failure based on the current NCE failure rate, as well as the probability of failure assuming that all NCEs still under development are successful (i.e., a more optimistic probability of failure). Our pessimistic prior probability model refers to the data that are based on the current failure rate, whereas the optimistic prior probability model refers to the failure data that are based on the assumption that all NCEs still under development will not fail.

For each prior probability model we computed the posterior probability of clinical success for each of the 18 NCE-indication pairs in our independent data set. The NCE outcome definitions were structured to create three sets of binary outcomes (success vs. failure). The rationale for using three sets of binary outcomes is that we aimed to assess our model in the setting of all possible assumptions regarding the outcomes for NCEs still under development. Therefore, we categorized our outcome data in three ways: a) active phase III/marketed vs. phase III terminated/pre-phase III, in which an NCE is deemed successful if it is either still in active phase III development or has received NDA approval; b) marketed vs. not marketed, in which an NCE is successful only if it has received NDA approval; and c) optimistic, for which we assumed that all active NCE development programs, regardless of latest phase of activity, would result in NDA approval. Each binary outcome definition classified all 18 NCEs as successful or failed, resulting in three lists of outcomes for the 18 NCEs (one list for each outcome group). We used the pessimistic prior probabilities to evaluate outcome groups (a) and (b), and we used the optimistic posterior probabilities to evaluate outcome group (c). We recorded the resultant posterior probability expectations for each NCE-indication pair. We then plotted receiver operating characteristic (ROC) curves, and calculated the c-index [17] to determine the area under the curve (AUC) for each ROC curve. We determined the optimum sensitivity and specificity values for each ROC curve.

Pharmacoeconomic Evaluation

We evaluated the economic impact the model would have if used by drug development teams, given the

measured sensitivities and specificities of the three outcome models on the independent data set, as well as the sensitivity and specificity of the three outcome models combined. Our goal was to compare our BN model's sensitivity and specificity values with the pharmaceutical industry's sensitivity and specificity values by subjecting each model's values to the Monte Carlo simulation, sampling from reported distributions for capitalized phase costs and cumulative 7-year revenues.

We used reported pharmaceutical industry performance data [1] to estimate the pharmaceutical industry's sensitivity and specificity for predicting success of phase II NCEs through phase III to NDA approval as follows: The probability that a phase II NCE will enter phase III is 31.4% [1], and 68.5% of phase III NCEs received NDA approval [1]. The proportion of true positive (TP) NCEs (successful NCEs) is therefore $0.314 \times 0.685 = 0.215$. The proportion of false positive (FP) NCEs (proceeded through phase III but failed to get NDA approval) is $0.314 - 0.215 = 0.099$. We were unable to determine the pharmaceutical industry's true negative (TN) and false negative (FN) rates, because it is impossible to determine what the fate of terminated NCEs would have been. It is likely that a proportion of terminated NCEs may have ultimately proven to be successful. The rationale for including TN and FN in our evaluation is that 1) TN and particularly FN have a significant impact on NCE costs and revenues because terminated would-be successful NCEs result in development costs and lost potential revenue; and 2) both TN and FN are required to estimate the pharmaceutical industry's sensitivities and specificities for predicting NCE success. Therefore, we computed the remaining proportion representing the sum of the pharmaceutical industry's TN and FN, as $1 - (TP + FP) = 0.686$. We constructed a flat distribution of the range from 0 to 0.686, with 5000 bins, to include in the Monte Carlo simulation. We selected a flat distribution for the range of possible TN and FN values for the pharmaceutical industry because there are no data regarding the would-be outcomes of terminated NCEs and we therefore did not want to bias the simulation by emphasizing one or more regions of the distribution. Each Monte Carlo simulation consisted of 2000 cycles. Individual data values are sampled from each distribution within each cycle of the Monte Carlo simulation.

The overall goal of the Monte Carlo simulations is to determine the expenditures and revenues for our BN model and for the pharmaceutical industry, to permit economically oriented comparisons. We submitted the pharmaceutical model to the Monte Carlo simulation three times to ensure the same number of simulation cycles as performed for the three combined BN outcome models. Each of the three pharmaceutical industry models underwent two consecutive Monte Carlo simulations. The role of the first simulation was solely to determine the sensitivity and specificity over 2000 cycles, sampling from the flat distribution of FN and TN in each simulation cycle, and computing each cycle's sensitivity (TP/[TP + FN]) and specificity (TN/[TN + FP]). The second Monte Carlo simulation to determine costs and revenues proceeded as for our BN models, described below.

To express all monetary data in year 2000 dollars, we used the current and real gross domestic product (GDP) data [18] to calculate the GDP Implicit Price Deflator, to deflate cost figures derived from data [19] expressed in 1992 dollars. We used published data [1,19] to set the cost ranges and distributions for the Monte Carlo simulations. Capitalized phase costs [1] used in the Monte Carlo simulation were (in millions of year 2000 dollars) \$30.5, \$41.6, and \$119.2, for phase I, phase II, and phase III, respectively. Phase costs were expressed as normal distributions, based on the reported normal distribution of total capitalized costs [1] (95% confidence limits \$684–936 million around a mean of \$802 million).

Revenue range for approved NCEs was estimated from 7-year postapproval cumulative mean (\$600 million) and top-decile (\$3095 million) values [19]. The revenue parameter values were randomly sampled from a gamma distribution, with a shape approximating the distribution of the reported return on investment for 18 pharmaceutical companies [19].

For each cycle of the Monte Carlo simulation, we calculated the expenditures per successful virtual NCE, and 7-year cumulative revenues per phase III trial given the model's submitted sensitivity and specificity values. We performed Monte Carlo simulation, randomly sampling from the reported cost and revenue distributions. All distributions were divided into 5000 bins, and each Monte Carlo simulation ran for 2000 cycles. Identical simulation parameters were used to evaluate each of the pairs of sensitivity and specificity values for our BN model and for the pharmaceutical industry performances.

Within each cycle of the Monte Carlo simulation, expenditures per successful NCE, and 7-year cumulative revenues per phase III trial were calculated given: 1) the model's sensitivity and specificity for predicting clinical success; 2) the phase cost values randomly sampled from the normally distributed cost distributions; and 3) the revenue value randomly sampled from the gamma-distributed revenues. Each cycle used a virtual group of 100 NCEs for which the submitted model's sensitivity and specificity values were used to determine the proportion of the 100 virtual NCEs that were accurately classified, which in turn permitted expenditure and revenue calculations. We assigned the virtual group of 100 NCEs an inherent success rate of 35% (i.e., 35 phase II NCEs would achieve NDA approval in the virtual group of 100 NCEs). We set this

Table 1 Bayesian network (BN) model performances for the three outcomes

Outcome	Active phase III/marketed	Marketed	Optimistic	Combined
Sensitivity (%)	89	67	75	80
Specificity (%)	89	83	64	76
Accuracy (%)	89	75	70	78
ROC AUC	0.83	0.61	0.68	0.72

AUC, area under the curve; ROC, receiver operating characteristic curve.

inherent success rate to exceed reported success rates for phase II NCEs [1] which is 0.215, the product of the probability that a phase II NCE will enter phase III (0.314) and the probability that a phase III NCE will achieve NDA approval (0.685). The virtual group of 100 NCEs must have a success rate exceeding reported pharmaceutical industry success rates to ensure that the possibility of FN NCEs would be included in the simulation. Changing the inherent success rate of the virtual group of 100 NCEs had no impact on the resultant differences between our BN model and the pharmaceutical industry model.

Using the sensitivity and specificity values submitted to the simulator, the proportion of TP, FP, TN, and FN NCEs was calculated for the virtual group of 100 NCEs. The expenditure calculation summed the sampled phase I and phase II costs for all 100 NCEs, added the phase III costs for the proportion of NCEs that the model correctly (TP) or incorrectly (FP) predicted to be successful, and divided the sum by TP to yield the cost per successful NCE. The revenue calculation multiplied TP by the sampled 7-year revenue value, and divided the product by the sum of TP and FP (i.e., all phase II NCEs selected for phase III trials), to yield the cumulative 7-year revenues per phase III trial.

We compared expenditures and revenues between our BN model and the pharmaceutical industry. We used the *t*-test for normally distributed data (expenditures) and the Mann–Whitney test for nonparametric data (revenues). The Bonferroni correction factor was applied to account for multiple tests. Monte Carlo simulation was executed within the R programming environment [20].

Results

Independent Data Set Characteristics

Publicly accessible pre-phase III data were accrued for 14 NCEs common to the ASCO and PharmaProjects data sets, with 18 NCE-indication pairs, of which nine were indication-consistent across in vitro, in vivo, and early human studies, in that all study environments were performed within the same indication. Latest-phase analysis revealed that a total of six NCEs received marketing approval. An additional two more NCEs were active in phase III development. Two NCEs were discontinued in phase III. Three NCEs were active in phase II and one NCE was discontinued in phase II.

Model Performance

Our BN model predicted the outcome on the independent test set with an optimal accuracy of 78% (80% sensitivity and 76% specificity). The sensitivities, specificities, accuracies, and ROC AUC values for the active phase III/marketed model, the marketed model, the optimistic model, and the combined performance of all three models are shown in Table 1. The ROC curves are shown in Figure 2.

Pharmacoeconomic Evaluation

Our BN model significantly reduced median expenditures per successful NCE by $283 million below pharmaceutical industry expenditures (39% reduction, from $727 million to $444 million, $P < 0.001$, Fig. 3, left). The BN model also significantly increased median cumulative 7-year revenues per phase III trial by $160 million above pharmaceutical industry revenues (46% increase, from $347 million to $507 million, $P < 0.001$, Fig. 3, right).

Comparisons of each of our BN's three outcome models individually against pharmaceutical industry performance revealed similar trends across all of the three outcome models (Table 2 and Fig. 4), with significant decreases in expenditures per successful NCEs (Fig. 4, top), and increases in revenues per phase III trial (Fig. 4, bottom).

Discussion

The impact of safety and toxicity on NCE failure is significant [9]. The cost of an NCE that will ultimately fail is directly proportional to the length of time between IND approval and termination of development. It follows that earlier and accurate termination of NCEs destined for failure results in: 1) significant development cost savings; 2) improved revenues because of due to selective development of successful NCEs; 3) freeing up clinical trial resources for other more promising agents in the development pipeline; and 4) limiting patient exposure to potentially unsafe and/or ineffective investigational agents. Nevertheless,

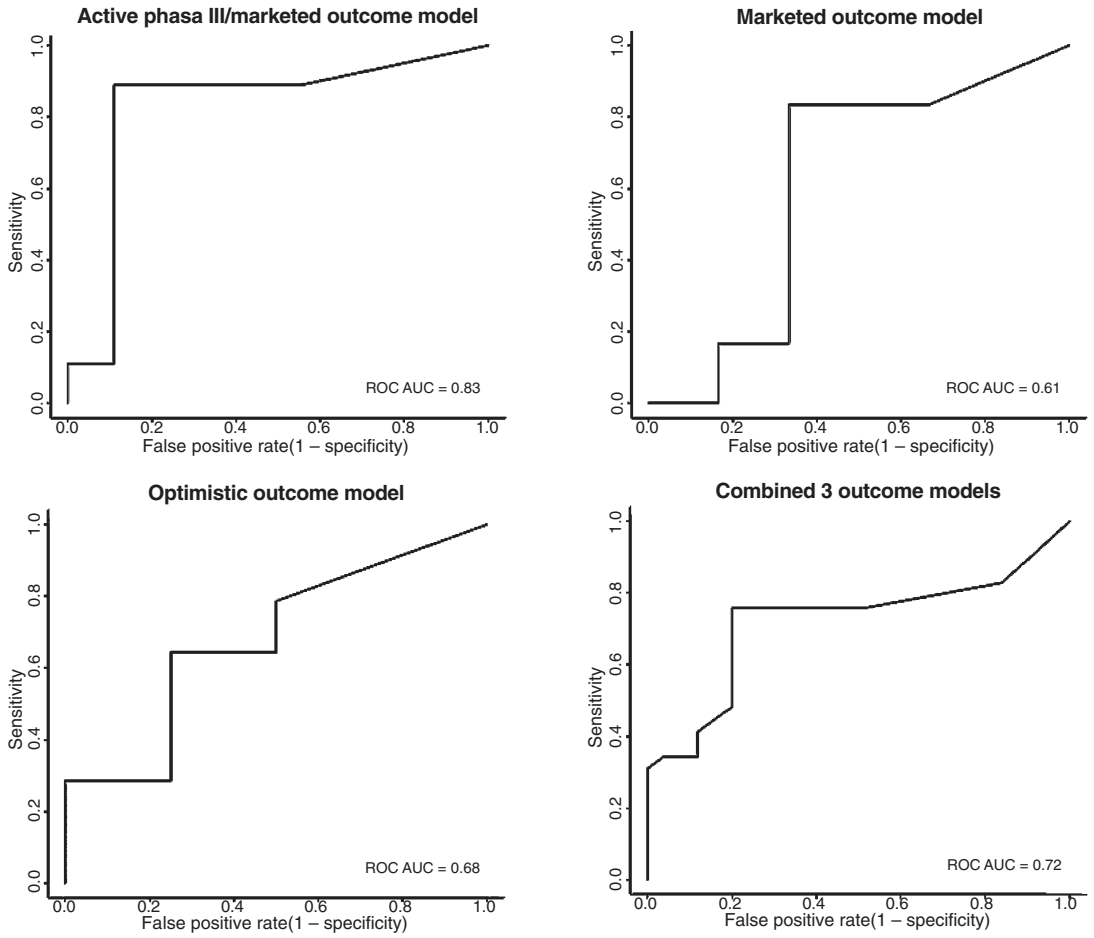

Figure 2 Receiver operating characteristic (ROC) curves for the Bayesian network models. AUC, area under the curve.

overzealous termination of NCEs will impede the development of novel effective therapies.

We have used publicly available data to construct and independently validate a BN model of drug development success. Based on our model's overall performance and economic impact, we have shown that go/no-go decisions that take prior data into account can result in significant savings and improved revenues. Terminating NCEs that would otherwise be late-stage failures would not only prevent higher expenditures but would also reduce exposing trial subjects to unsafe and ineffective NCEs, and would increase the proportion of pipeline NCEs that are ultimately successful clinically and financially. Ideally, these added revenues would be translated into lower pricing for brand name medications, thereby reducing the cost burden on patients and third party payers.

We can infer from our findings that the pharmaceutical industry's accuracy for predicting late-phase success is suboptimal. Even the modest predictive performances of our BN's marketed and optimistic models improve on current pharmaceutical industry performances. DiMasi et al. reported that, in comparing year 2000 analyses [1] with prior analyses [8], a smaller proportion of pharmaceutical industry failures are occurring in phase III (12.6%, reduced from 17.1% reported in the earlier study). Nevertheless, a reduction in NCE approvals ensued [21]. Taken together, it is reasonable to hypothesize that the pharmaceutical industry may have overcorrected for rising development costs due to later-stage failures, and in so doing has suppressed the development of novel therapies. Our BN model has demonstrated the ability to counter this trend, by increasing success-bound phase III development activity, as evidenced by significantly increased revenue.

The main distinction between our model and previously described models is that our model focuses on predicting the outcome of a specific NCE. Other models [7–9,22–25] have taken a population-based analysis approach, yielding valuable data on overall success rates, but not truly addressing the termination decision for a single, specific NCE. Other Bayesian approaches described in the literature differ from our approach with respect to the domain to which Bayes theorem is applied. Published Bayesian approaches

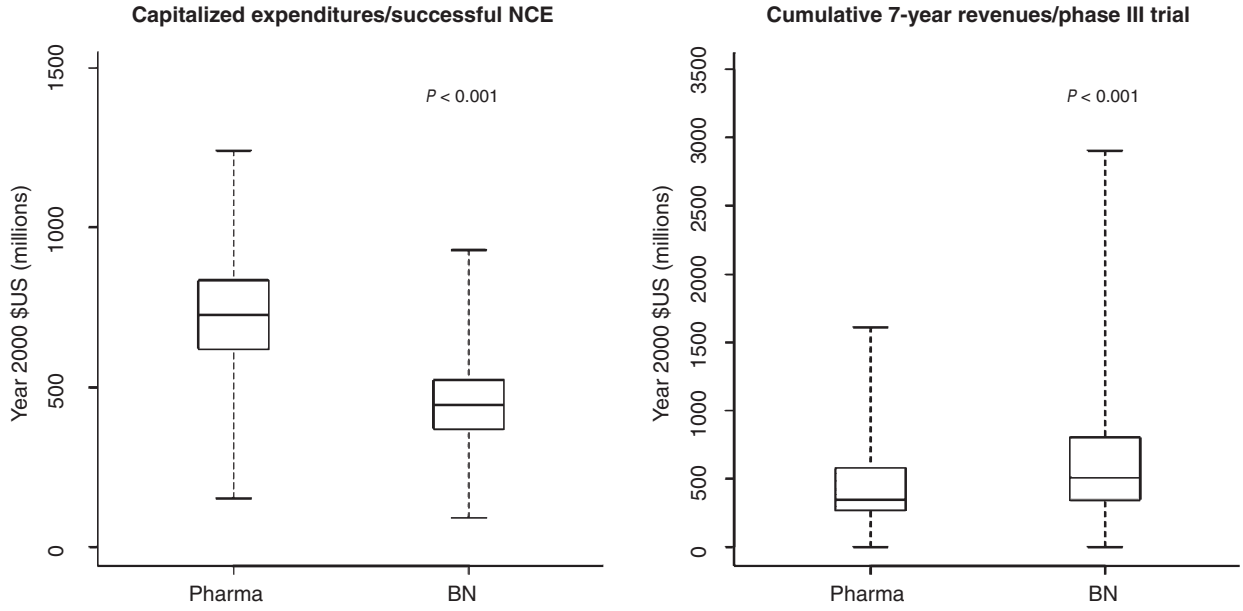

Figure 3 Box plots demonstrating capitalized expenditures per successful NCE (left) and cumulative 7-year revenues per phase III trial (right) for the pharmaceutical industry and our Bayesian network (BN) model (all three outcomes combined).

within the field of pharmaceutical development compare the benefits of Bayesian statistics over frequentist approaches and focus on the utilization of Bayesian statistics for the analysis of clinical trial data, which are in turn used to define "stopping boundaries" [26–29]. Berry et al. [27] used accumulated information on an NCE's performance to determine at which point a given clinical trial's evidence of efficacy is sufficiently negative that the trial should be discontinued, but did not perform sensitivity analyses, and did not address the issue of overall clinical success of the NCE. Recently, De Ridder [30] described a simulation approach to predict phase III outcomes based on phase II data for a single agent, and reported benefits with respect to trial design and dosing issues for the ongoing phase III trials of the drug in question. Nevertheless, given that only a single drug was modeled, no validation studies, sensitivity analyses, or pharmacoeconomic evaluation could be performed.

We acknowledge some limitations to our evaluation, mainly as a result of our inability to access proprietary data on terminated NCEs. We do not have sufficient data to evaluate NCE combinations, which are particularly common in the oncology domain, where monotherapy is rare. Similarly, we attempted to acquire indication-specific data, although this requirement further limited our ability to find complete data for monotherapy NCEs, and we therefore included some NCE data with "mixed" indications. Finding complete indication-specific, publicly accessible data for a given combination of NCEs was not possible. Ideally, our input data would be indication-consistent across all investigational environments, as well as consistent with the latest-stage outcome data. We were also further limited with respect to our outcome data, in that when categorizing the outcomes by active phase III/marketed, NCEs terminated in early phases are necessarily categorized as having a bad outcome, when in fact they may have been terminated appropriately. Conversely, NCEs still active in early phases are also categorized as having a bad outcome when the true outcomes for these NCEs are not yet known. For these reasons, we performed three evaluations to assess our model's performance assuming realistic (active phase

Table 2 Economic results for the pharmaceutical industry and the Bayesian Network (BN) models (median, interquartile range, in millions of year 2000 US dollars)

	Pharmaceutical industry	BN: All outcomes	BN: marketed/active phase III	BN: marketed	BN: optimistic
Expenditures per successful NCE	727 (620–836)	444 (370–524)	379 (324–438)	458 (389–528)	506 (432–580)
Revenues per phase III trial	347 (267–579)	507 (343–804)	626 (468–1039)	436 (334–755)	404 (310–653)
P-value (vs. pharmaceutical industry)	—	$P < 0.001$	$P < 0.001$	$P < 0.001$	$P < 0.001$

NCE, new chemical entity.

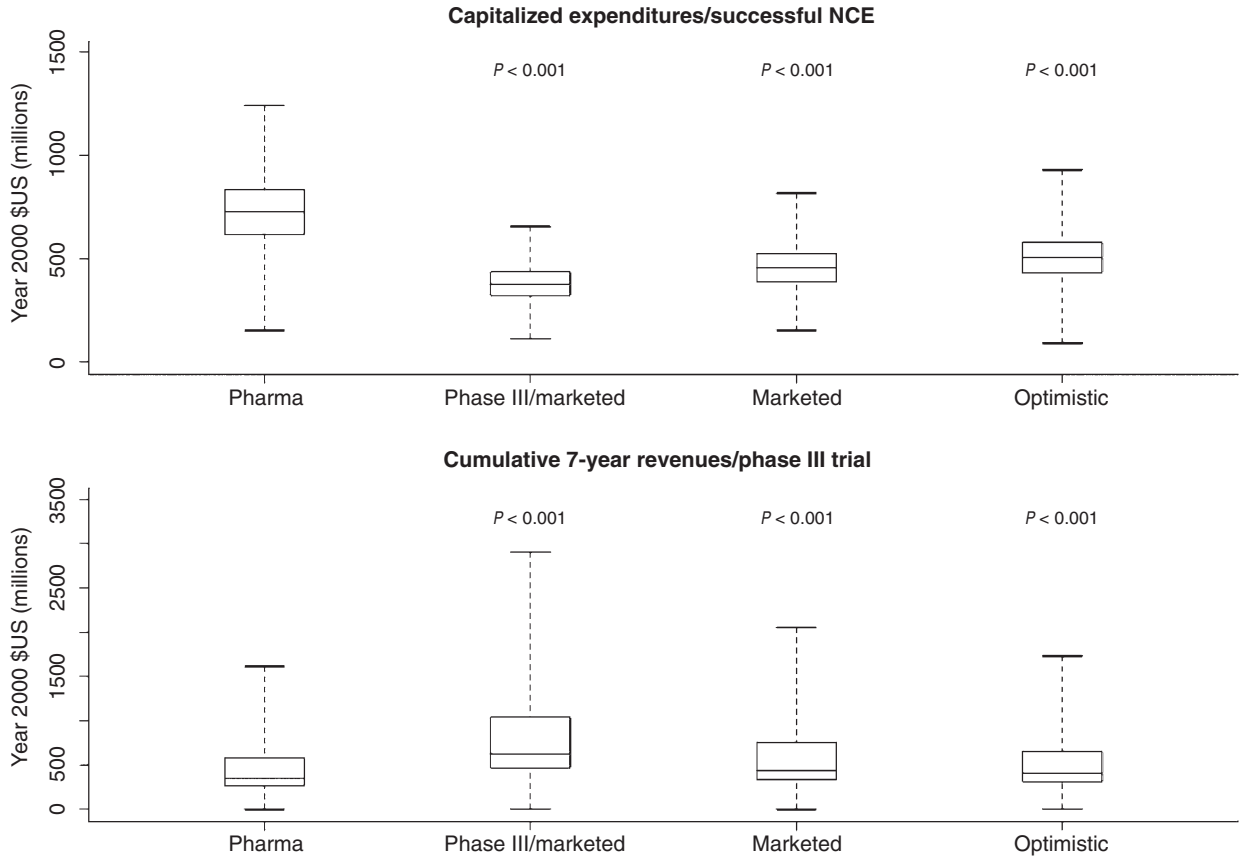

Figure 4 Box plots demonstrating capitalized expenditures per successful NCE (top) and cumulative 7-year revenues per phase III trial (bottom) for the pharmaceutical industry and each of the three outcome models of our Bayesian network (BN).

III/marketed), pessimistic (marketed only), and optimistic (all active will be marketed) outcomes. Ongoing efforts to construct mandatory clinical trial registries will improve access to independent validation data. Our independent data set may also be affected by publication bias, in that it is possible that negative studies are not published as frequently as promising data, although these biases, should they exist, have no bearing on the evaluation of our model. As well, our validation data set was confined to antineoplastic NCEs, and we generalized the resultant sensitivity and specificity values to determine pharmacoeconomic performance. Improved access to proprietary data for other therapeutic classes would be of interest for future evaluations.

In summary, NCE R&D expenditures are reportedly increasing, with a detrimental effect on the cost of living of millions of Americans. Our BN approach is novel in that individual NCE characteristics are modeled on the background of prior data specific to those same characteristics for similar NCEs. Despite limited access to NCE development data, our model demonstrated substantial improvement over reported pharmaceutical industry performances. Open access to proprietary, de-identified data from pharmaceutical companies, particularly data on terminated NCEs, would enhance our ability to further evaluate and optimize the performance of our model.

The data on which we evaluate our model may not be generalizable to pharmaceutical companies, but the structure of the model is directly applicable to use a given pharmaceutical or biotechnology company's stored data on previously failed compounds to improve the accuracy of go/no-go decisions for NCE's in the pipeline. We have shown that the use of prior data, perhaps even within-company data only, can potentially improve the predicted results of NCE's in the pipeline substantially. We hope that our report will encourage pharmaceutical companies to use their proprietary data for the benefit of their companies and for society in the form of safer and more effective pipeline drugs, and lower costs of development.

The authors wish to acknowledge Isaac S. Kohane, Robert H. Rubin, Peter Szolovits, and Stephen G. Pauker, for their guidance and assistance.

Sources of financial support: This work was supported in part by NIH Grant K23 RR16080 (ADS), and the Sloan

Foundation Industry Centers Fellowship (ADS). Parts of this work are described in USPTO patent application 20050021237.

References

1 DiMasi JA, Hansen RW, Grabowski HG. The price of innovation: new estimates of drug development costs. J Health Econ 2003;22:151–85.
2 PublicCitizen. Rx R&D Myths: The Case Against the Drug Industry's R&D 'Scare Card'. Congress Watch 2001.
3 Relman AS, Angell M. America's other drug problem. how the drug industry distorts medicine and politics. New Repub 2002;227:27–41.
4 Rawlins MD. Cutting the cost of drug development? Nat Rev Drug Discov 2004;3:360–4.
5 Riggs TL. Research and development costs for drugs. Lancet 2004;363:184.
6 Scherer FM. The pharmaceutical industry—prices and progress. N Engl J Med 2004;351:927–32.
7 DiMasi JA. Success rates for new drugs entering clinical testing in the United States. Clin Pharmacol Ther 1995;58:1–14.
8 DiMasi JA, Hansen RW, Grabowski HG, Lasagna L. Cost of innovation in the pharmaceutical industry. J Health Econ 1991;10:107–42.
9 DiMasi JA. Risks in new drug development: approval success rates for investigational drugs. Clin Pharmacol Ther 2001;69:297–307.
10 DiMasi JA. The value of improving the productivity of the drug development process: faster times and better decisions. Pharmacoeconomics 2002;20(Suppl. 3):S1–10.
11 Schachter AD. Probabilistic Modeling of the Drug Development Domain: A Bayesian Domain-Knowledge Application for Pharmacovigilance [Masters of Science]. Cambridge, MA: Massachusetts Institute of Technology, 2003.
12 Schachter AD, Ramoni MF Inventors. Method and apparatus for evaluating new chemical entities. USPTO patent application # 20050021237, USA, 2005.
13 Roberts TG Jr, Goulart BH, Squitieri L, et al. Trends in the risks and benefits to patients with cancer participating in phase 1 clinical trials. JAMA 2004;292:2130–40.
14 Pearl J. Fusion, propagation, and structuring in belief networks. Artif Intell 1986;29:241–88.
15 Pearl J. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. San Mateo, CA: Morgan Kaufmann, 1988.
16 Winston PH, Narasimhan S. On to Java (3rd ed.). Boston, MA: Addison-Wesley, 2001.
17 Dreiseitl S, Ohno-Machado L, Binder M. Comparing three-class diagnostic tests by three-way ROC analysis. Med Decis Making 2000;20:323–31.
18 Bureau of Economic Analysis. National Economic Accounts. In: USDOC, ed. Available from: http://www.bea.gov/bea/dn/home/gdp.htm [Accessed July 12, 2005].
19 Grabowski HG, Vernon J. The distribution of sales revenues from pharmaceutical innovation. Pharmacoeconomics 2000;18(Suppl. 1):S21–32.
20 R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, 2005. Accessed from: http://www.r-project.org [Accessed June 17, 2005].
21 Frantz S, Smith A. New drug approvals for 2002. Nat Rev Drug Discov 2003;2:95–6.
22 Sheck L, Cox C, Davis HT, et al. Success rates in the United States drug development system. Clin Pharmacol Ther 1984;36:574–83.
23 Bienz-Tadmor B, Dicerbo PA, Tadmor G, Lasagna L. Biopharmaceuticals and conventional drugs: clinical success rates. Biotechnology (N Y) 1992;10:521–5.
24 Struck MM. Biopharmaceutical R&D success rates and development times. A new analysis provides benchmarks for the future. Biotechnology (N Y) 1994;12:674–7.
25 DiMasi JA, Hansen RW, Grabowski HG, Lasagna L. Research and development costs for new drugs by therapeutic category. A study of the US pharmaceutical industry. Pharmacoeconomics 1995;7:152–69.
26 Berry DA. Interim analyses in clinical trials: classical vs. Bayesian approaches. Stat Med 1985;4:521–6.
27 Berry DA, Ho CH. One-sided sequential stopping boundaries for clinical trials: a decision-theoretic approach. Biometrics 1988;44:219–27.
28 Berry DA. A case for Bayesianism in clinical trials. Stat Med 1993;12:1377–93; Discussion 1395–404.
29 Spiegelhalter DJ, Freedman LS, Parmar MKB. Applying Bayesian ideas in drug development and clinical trials. Stat Med 1993;12:1501–11.
30 De Ridder F. Predicting the outcome of phase III trials using phase II data: a case study of clinical trial simulation in late stage drug development. Basic Clin Pharmacol Toxicol 2005;96:235–41.

Supplementary material for this article can be found at: http://www.ispor.org/publications/value/ViHsupplementary.asp