Congressional Budget Office
Nonpartisan Analysis for the U.S. Congress

CBO's Model of Drug Development: Ongoing Updates

April 4, 2025

Presentation at the American Enterprise Institute

Christopher Adams

Health Analysis Division

For information about the event, see https://tinyurl.com/ysvx7xfe.

CBO's Drug Development Model

The Congressional Budget Office's model is intended to help the Congress understand how legislative proposals would affect the development of new drugs.

The model allows the agency to provide information about drug development with its estimates of the budgetary costs of legislation.

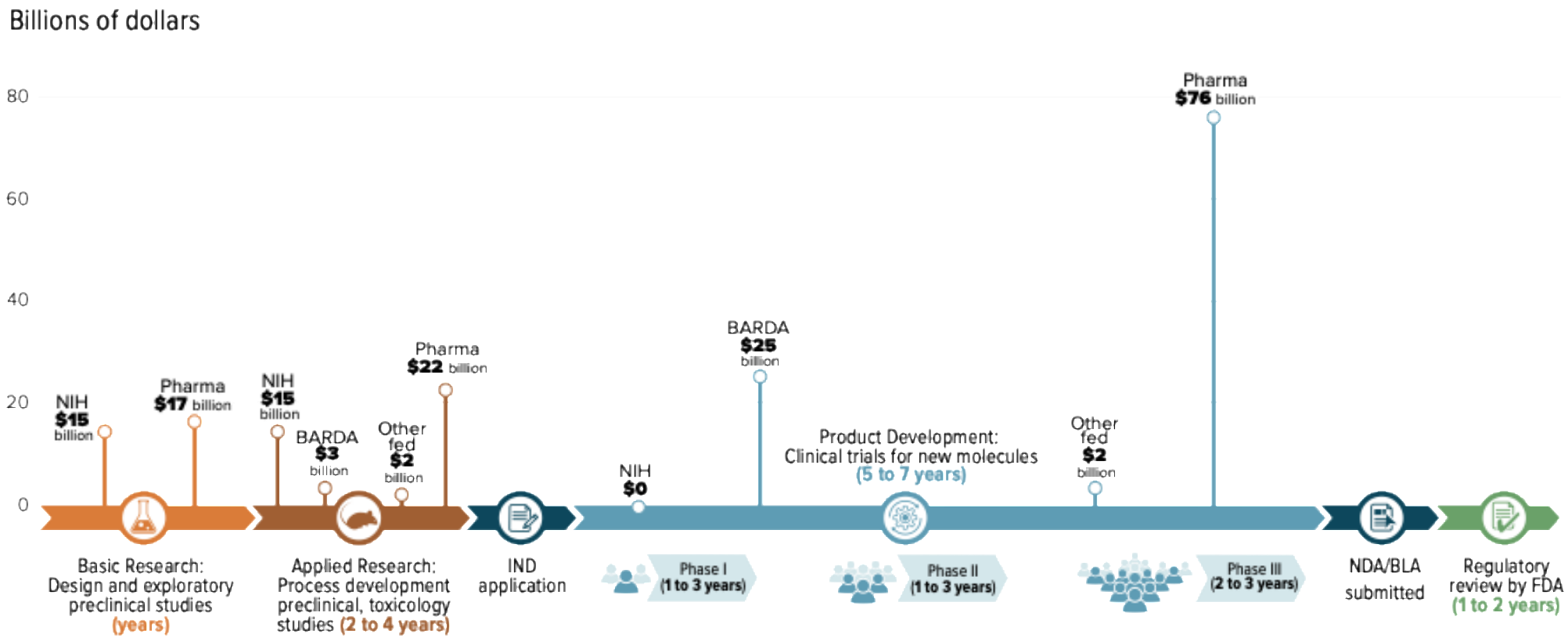
CBO is updating its drug development model to address feedback from the Congress, responses to CBO's December 2023 "Call for New Research" blog post, and other input. The agency continues to seek input to improve the model.

This update constitutes a significant advance in the agency's ability to analyze the development of new drugs—by now accounting for correlation across phases of drug development and differences in drugs' characteristics.

Congressional Budget Office, "A Call for New Research in the Area of New Drug Development," *CBO Blog* (December 20, 2023), www.cbo.gov/publication/59818.

Background

Government and Private Sector Funding of Biomedical Research and Development, 2022



Sources: National Science Foundation, “Survey of Federal Funds for Research and Development, 2022–2023,” https://tinyurl.com/4c8bha6d, and “Business Enterprise Research and Development (BERD) Survey,” https://tinyurl.com/7h6weuks. BARDA = Biomedical Advanced Research and Development Authority; BLA = biologics license application; FDA = Food and Drug Administration; IND = investigational new drug; NDA = new drug application.

Overview of CBO's Drug Development Model

CBO models pharmaceutical firms' decision-making over time about whether to move a given drug through the development process.

To bring a drug to market, a firm is assumed to have four decision points: phase 0 (preclinical studies), phase I, phase II, and phase III.

CBO treats phase 0, which is outside the model, as exogenous. The agency is updating the modeling of decision-making in phases I, II, and III.

The firm analyzes expected returns and expected costs for its drug candidate. If expected returns are greater than expected costs, the firm chooses to enter the development phase. The firm accounts for the option value (the value of maintaining the ability to continue to develop the drug).

CBO estimates the parameters of its model using drug-level data on net revenues, development expenditures, and time in development, as well as drug-level rates of success and entry for each phase of development.

Major Updates to CBO's Drug Development Model

CBO has continued to update its drug development model since the agency estimated the budgetary effects of the prescription drug provisions in the 2022 reconciliation act.

The updated model incorporates **new data** that either were not previously available or are more granular, including drug-level data on entry into each phase of development and total U.S. revenues.

The new model **jointly estimates firms' decision-making at each stage** of drug development, allowing for signals of a drug's revenues and costs to be correlated across decision points. The previous model produced estimates that did not account for such correlation.

The updated model uses a **regression to account for variation** by a drug's characteristics and allows CBO to estimate the effects of various policy changes on the development of different types of drugs. The previous model, by contrast, provided a single aggregate estimate for all drugs.

Presentation Outline

- The Model
- Updated Sources of Data
- Updated Model Estimation
- Next Steps and Areas Where Additional Research Would Be Helpful

The Model

Overview of Decision-Making Model

CBO's model of drug development is a dynamic model of firms' decision-making to advance a drug candidate through the development phases. The model is solved through backward induction.

The firm decides before each phase (I, II, III) whether to move forward. The decision to enter later phases depends on the successful completion of the earlier phase(s). For example, the firm can consider entering phase II if it has successfully completed phase I. ("Successful" completion of phases I and II means only that the trials were completed; it does not imply anything about the results of those trials.)

The firm's decision at each point is based on a draw of expected revenues and expected costs, which the model allows to be correlated across decision points. The firm's decision accounts for the option value of entering earlier phases of development (I and II).

In deciding whether to move a drug candidate forward, the firm also considers the drug's characteristics (including the number of competitors in the class, the route of administration, and whether the drug is a biologic), since those features can affect expected costs and returns.

Decision-Making Model at Each Development Phase

Phase III entry decision

- Profit calculation: Discounted present value revenues x probability of success − capitalized cost
- Conditional on entry to phases I and II
- Estimated on a set of drugs that have successfully completed at least one phase II clinical trial

Phase II entry decision

- Profit calculation: Expected net discounted returns x probability of success − capitalized cost
- Conditional on entry to phase I
- Estimated on a set of drugs that have successfully completed at least one phase I clinical trial

Phase I entry decision

- Profit calculation: Expected net discounted returns x probability of success − capitalized cost
- Estimated on a set of drugs for which preclinical studies have been conducted

Expected net discounted returns for a given phase of development equal the greater of expected discounted revenues minus capitalized development costs in later phases or zero.

Overview of Decision-Making Model

Let $k \in \{1, \ldots, \mathrm{K}\}$ represent a set of drug candidates.

Draw $\{R_k, C_{3k}, C_{2k}, C_{1k}\} \sim F$.

- Phase III entry decision, $(\pi_{3k} = 1)$: $\mathrm{p}_{3k} R_k > C_{3k}$ conditional on $\pi_{2k} = 1$ and $\pi_{1k} = 1$
- Phase II entry decision, $(\pi_{2k} = 1)$: $\mathrm{p}_{2k} \mathbb{E}(\max\{\mathrm{p}_{3k} R_k - C_{3k}, 0\} | C_{2k}) > C_{2k}$ conditional on $\pi_{1k} = 1$
- Phase I entry decision, $(\pi_{1k} = 1)$: $\mathrm{p}_{1k} \mathbb{E}\,(\max\{\mathrm{p}_{2k} \mathbb{E}(\max\{\mathrm{p}_{3k} R_k - C_{3k}, 0\} | C_{2k}) - C_{2k}, 0\} | C_{1k}) > C_{1k}$

In this model, CBO uses data on drug revenues, costs for each phase of development, and probabilities of entry into and successful completion of each phase of development.

R_k is a draw from a distribution of revenues, and C_{3k}, C_{2k}, C_{1k} are draws from a distribution of the costs of each phase of development for drug *k*. Revenues and costs are jointly distributed according to distribution F. p_{1k}, p_{2k} , p_{3k} are the probability of success in each phase for drug *k*. π_{1k} and π_{2k} are indicators that drug *k* successfully completed previous phases of development.

Updated Sources of Data

Data Sources in the Updated Model

Model component	Data source for original model	Data source for updated model	Advantage of updated data
Drug revenues	Medicare Part D	SSR Health	Total U.S. revenues rather than revenues in Part D only
Development costs (phases I, II, and III)	DiMasi, Grabowski, and Hansen (2016)	Wouters, McKee, and Luyten (2020)	Drug-level reported expenditures and time in development rather than survey data
Phase success probabilities	DiMasi, Grabowski, and Hansen (2016)	Cortellis database and ClinicalTrials.gov	Drug-level success rates for each phase of development rather than survey data
Phase entry probabilities	Calibration	Cortellis database and ClinicalTrials.gov	Drug-level data on entry into each phase of development

Olivier J. Wouters, Martin McKee, and Jeroen Luyten, "Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009–2018," *Journal of the American Medical Association*, vol. 323, no. 9 (March 2020), pp. 844–853, https://doi.org/10.1001/jama.2020.1166; and Joseph A. DiMasi, Henry G. Grabowski , and Ronald W. Hansen, "Innovation in the Pharmaceutical Industry: New Estimates of R&D Costs," *Journal of Health Economics*, vol. 47 (May 2016), pp. 20–33, https://doi.org/10.1016/j.jhealeco.2016.01.012.

Overview of Predictions in the Updated Model

Old model	New model
For **all** drug candidates in the aggregate, without regard to individual drugs' characteristics, estimate the distribution of the following: ▪ Lifetime revenues ▪ Clinical trial costs for each phase of development ▪ Trial success rates for each phase of development	For **each** drug candidate: ▪ If data for revenues or phase costs are available, match those data to the drug candidate. ▪ If data are missing, predict revenues or phase costs using the drug's observed characteristics, such as therapeutic class and route of administration. ▪ Predict success rates for clinical trials using the drug candidate's observed characteristics.

How CBO Estimates Drug-Level Revenues

CBO uses drug-level data from SSR Health for total net revenues in the United States for brand-name drugs from 2008 to 2023.

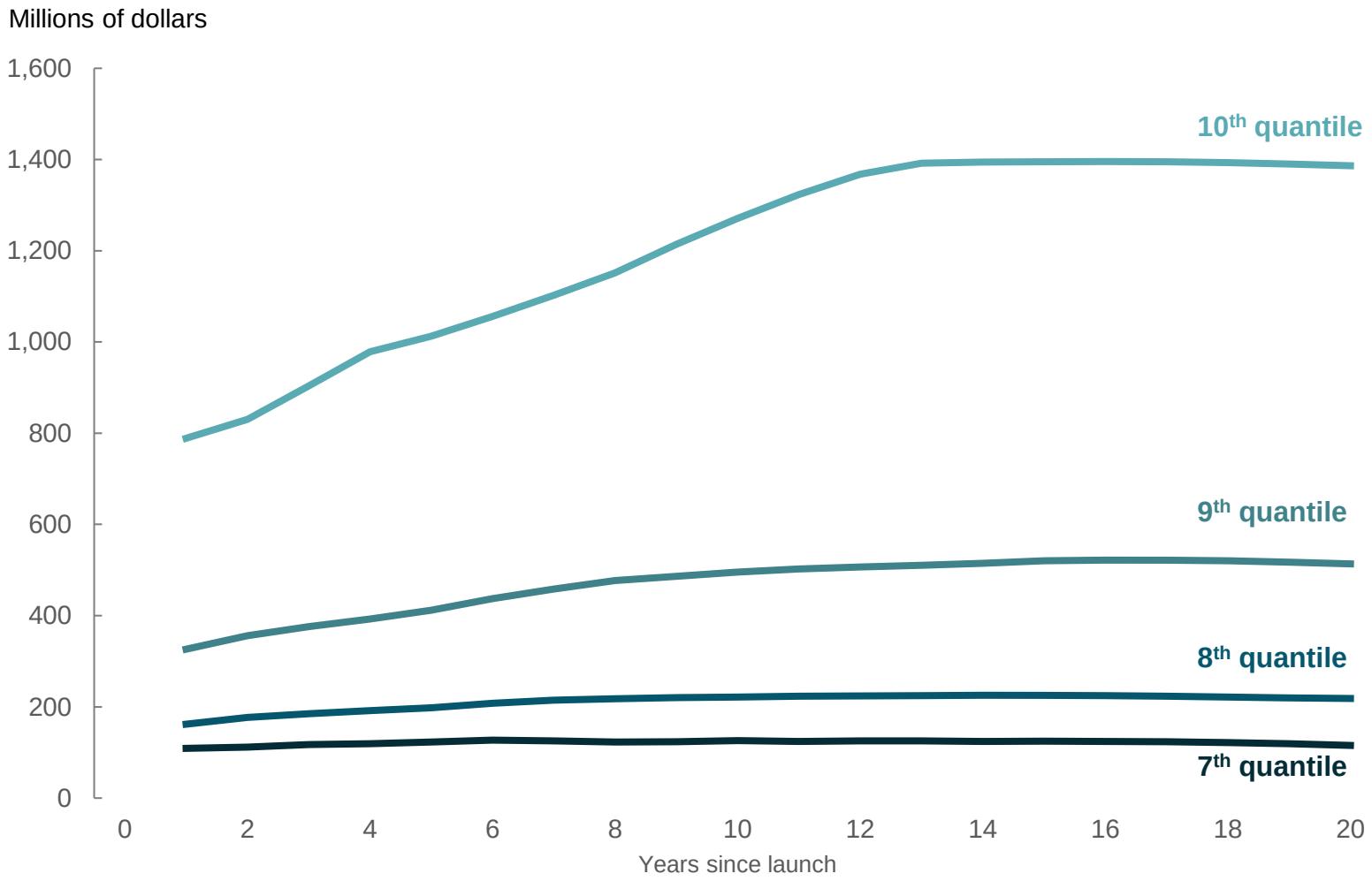
The agency computes lifetime revenues for drugs on the market by summing annual revenues. In cases in which the annual data are missing, the agency imputes the annual number on the basis of the drug's observed characteristics and years since launch. CBO uses a net present value based on a discount rate of 0.086 (Damodaran).

For each drug candidate, CBO uses estimates of lifetime revenues. In instances in which those data are unavailable, the agency predicts lifetime revenues on the basis of observed characteristics of the drug, such as therapeutic class and route of administration.

CBO's current estimates of lifetime drug-level revenues focus on those in the United States but may expand to include such revenues worldwide.

Aswath Damodaran, "Costs of Capital by Industry" (accessed March 18, 2025), https://tinyurl.com/32z4p68s.

Median Estimated Annual Drug-Level Revenues, by Years Since Launch



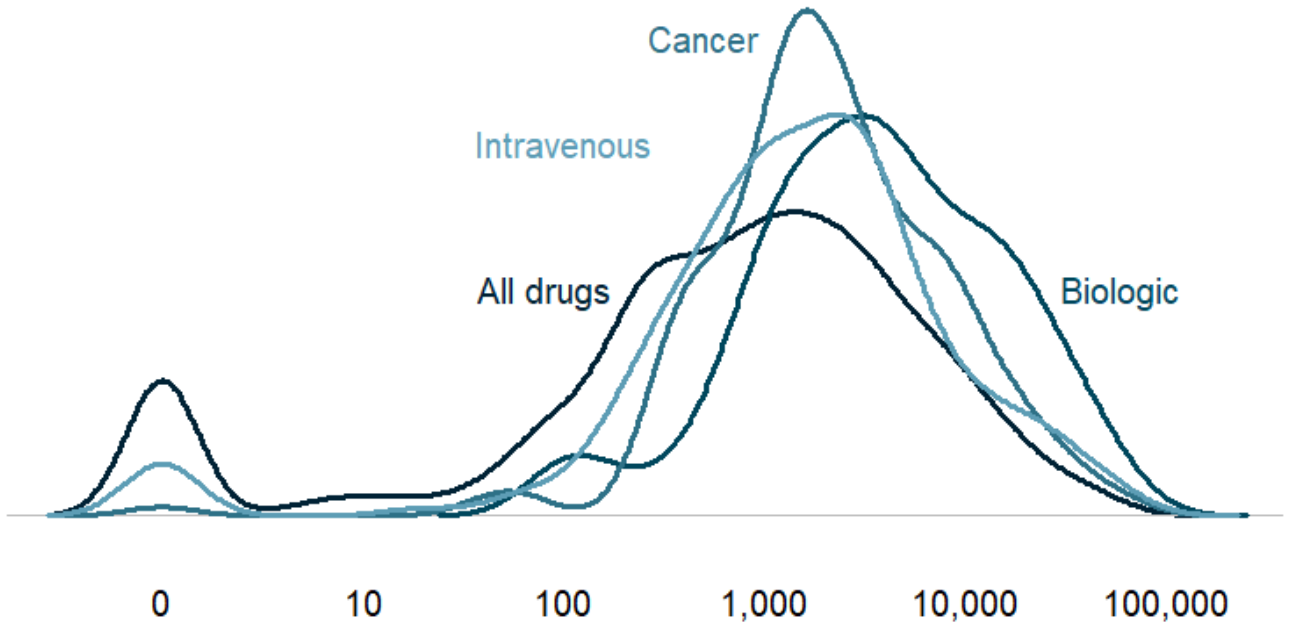
Drugs are categorized into 10 groups (quantiles) by estimated lifetime revenue. The figure shows median annual revenue for drugs in the top four categories.

For drugs in the 7th and 8th quantiles, median revenue increases each year for the first 10 to 15 years. It then starts to decline.

For drugs in the 9th and 10th quantiles, median revenue increases for at least the first 15 years. Median revenue starts to fall between 15 and 20 years after launch.

CBO is currently investigating the revenue trajectory of drugs launched more than 20 years ago.

Lifetime Discounted Revenues, by Drug Type



The updated methodology allows CBO to predict lifetime discounted revenues by drug type.

Over their lifetimes, biologic drugs earn substantially more revenues than small molecule drugs. Cancer and intravenous (IV) drugs earn somewhat more than drugs overall, partly because there are fewer low-revenue drugs in those categories.

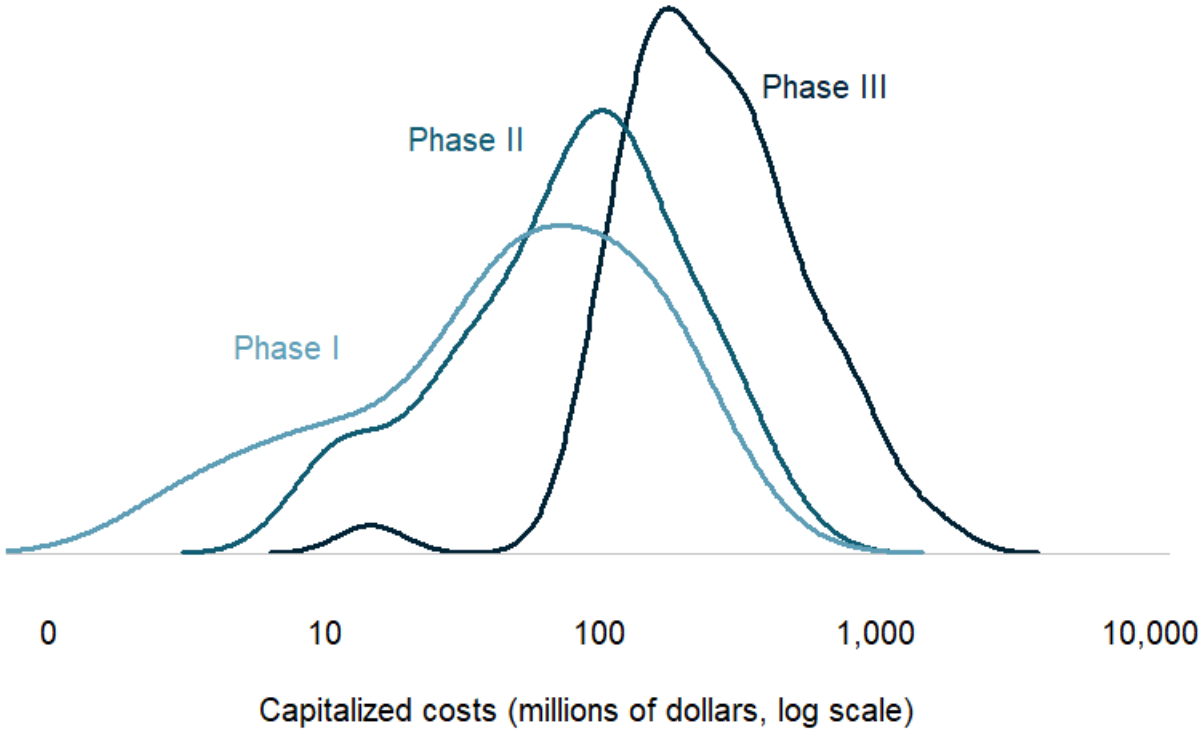
How CBO Estimates Drug-Level Development Costs

CBO uses data from Wouters, McKee, and Luyten (2020) on drug-specific development expenditures by phase and time in development for a sample of drugs. Those data come from public financial reports and include information from 2009 through 2018.

CBO capitalizes development expenditures forward to the time of market launch using a discount rate of 0.086. The agency assumes that spending is evenly distributed across time during a given phase.

In instances in which CBO does not have data for capitalized costs, the agency estimates costs on the basis of the drug's observed characteristics in the Wouters, McKee, and Luyten (2020) data, such as therapeutic class and route of administration.

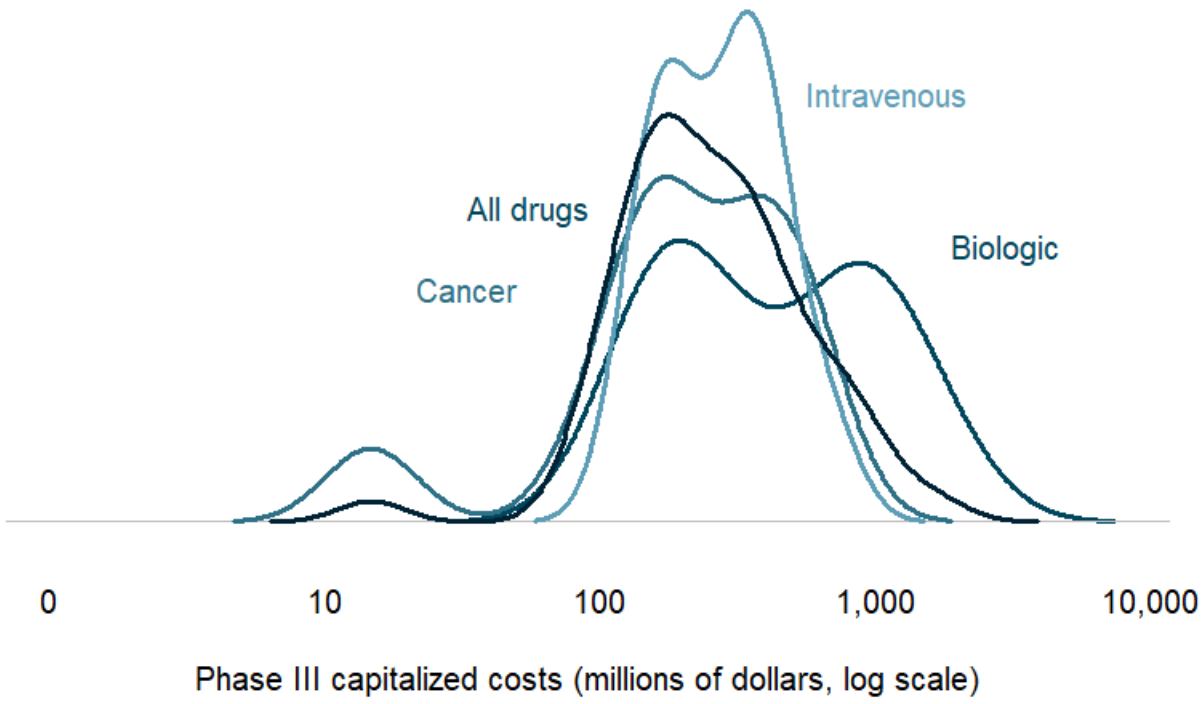
Drug-Level Development Costs, by Phase of Development



The updated methodology allows CBO to estimate the distribution of capitalized costs by phase of development.

Phase III is substantially more costly than phase II, which is more costly than phase I.

Phase III Development Costs, by Drug Type



CBO can estimate the distribution of capitalized costs for phase III by broad categories of drugs using data from Wouters, McKee, and Luyten (2020) and a discount rate of 0.086.

Biologics are more expensive to bring through phase III. Intravenous drugs have high development costs, but development costs vary less for that category of drugs than they do for all drugs or for all biologic drugs.

Updated Data on Successful Completion of Each Development Phase

CBO links data from Cortellis and Clinicaltrials.gov from 2000 to 2022. Those data provide information about drugs as they move through phases of development.

Success in phase III is defined as entry on the market. Success in phases II and I is defined as successful completion of at least one human clinical trial in the given phase.

CBO treats the outcome of “successful completion” as exogenous; that is, once a firm decides to enter a phase of development, the firm may not be able to control whether that phase is successfully completed (for example, a clinical trial may not be successful because the trial cannot enroll enough patients).

CBO predicts the success rate of all drug candidates. To do that, the agency performs a regression analysis on observed characteristics of a drug, such as its therapeutic class and route of administration.

Variation in Success Rates, by Development Phase and Drugs' Characteristics

	Dependent variable:		
	Success		
	Phase III	Phase II	Phase I
Cancer	-0.056***	-0.319***	-0.400***
	(0.017)	(0.034)	(0.034)
Infection	0.053***	0.113***	0.139***
	(0.017)	(0.040)	(0.043)
Cardiovascular	0.026	0.130**	0.113*
	(0.021)	(0.053)	(0.062)
Biologic	0.030*	0.100**	0.041
	(0.017)	(0.039)	(0.040)
Oral	0.215***	0.079*	0.178***
	(0.017)	(0.041)	(0.044)
IV	0.126***	-0.096**	0.010
	(0.019)	(0.041)	(0.040)
Constant	0.303***	0.957***	1.104***
	(0.014)	(0.035)	(0.039)
Observations	4,920	9,062	9,925
Log likelihood	-3,399.777	-4,158.258	-4,032.456
Note:		*p<0.1; **p<0.05; ***p<0.01	

CBO uses a probit regression (in which outcomes are either the successful completion of the trial or its failure by not reaching completion) to estimate the likelihood that a given drug will successfully complete a certain phase of development.

Across most phases of development, cancer drugs have lower success rates and oral drugs have higher success rates.

Updated Model Estimation

Reduced-Form Analysis of the Relationship Between Costs and Revenues

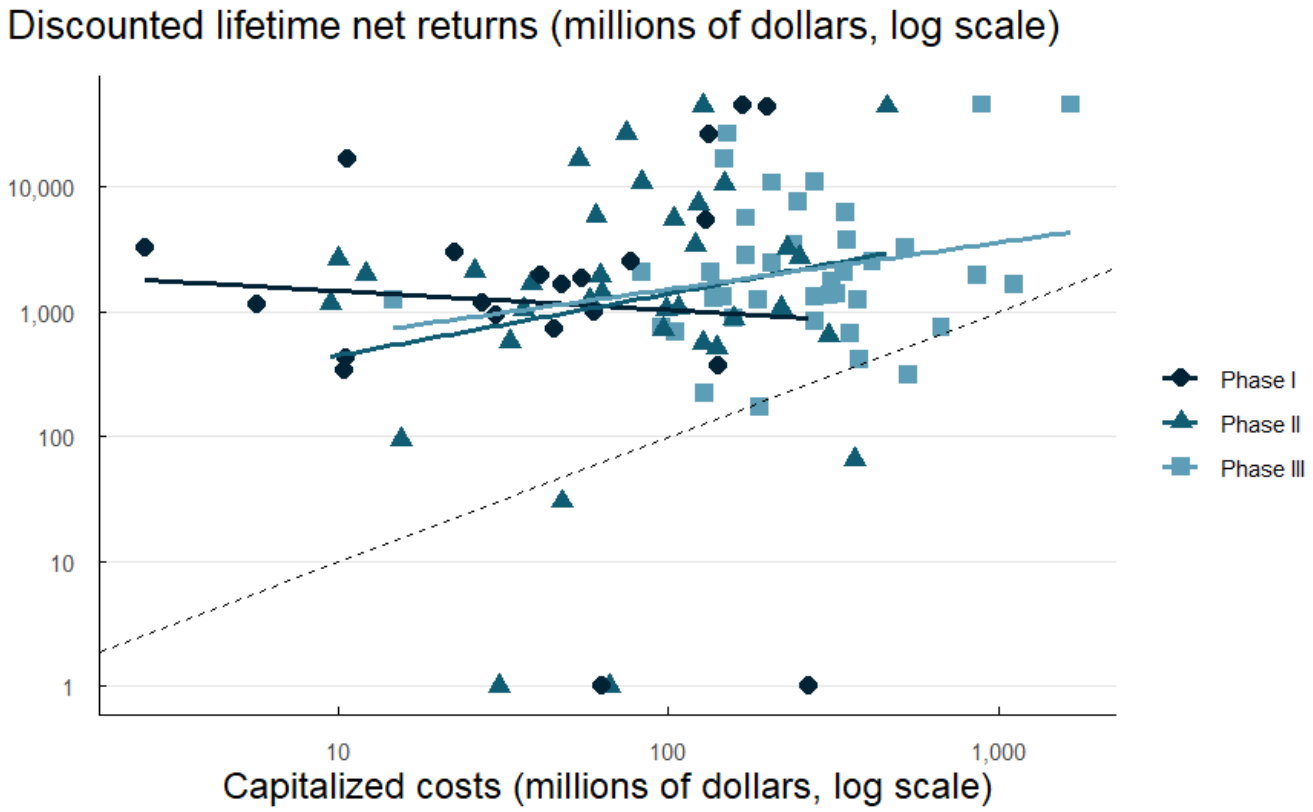
Model estimation involves matching development costs and lifetime discounted revenues by drug. Using the data discussed in the previous section, CBO can examine the drug-level relationship between costs and revenues outside the structural model.

CBO estimates the reduced-form relationship between development costs (from Wouters, McKee, and Luyten (2020)) and lifetime discounted revenues (from SSR Health) for each phase of development.

- The phase I return is calculated as the expected revenues minus the expected phase III cost minus the expected phase II cost.

- The phase II return is calculated as the expected revenues minus the expected phase III cost.

The Reduced-Form Relationship Between Costs and Returns



Each dot denotes a drug in a given phase of development.

The 45-degree dashed line denotes zero profit. As expected, most drugs are above that line, indicating higher returns than costs.

The solid lines represent the relationship between returns and costs. In later phases of development, drugs with higher costs typically also have higher returns.

Reduced-Form Analysis of Entry Probabilities for Each Development Phase

	Dependent variable:		
	Enter		
	Phase III	Phase II	Phase I
Cancer	0.023	-0.157***	0.010
	(0.063)	(0.053)	(0.029)
Infection	0.158***	0.123**	0.108***
	(0.061)	(0.056)	(0.038)
Cardiovascular	0.161**	0.358***	0.064
	(0.077)	(0.082)	(0.062)
Biologic	0.035	0.129**	0.358***
	(0.063)	(0.058)	(0.030)
Oral	0.102*	0.326***	1.139***
	(0.061)	(0.061)	(0.036)
IV	0.157**	0.219***	1.212***
	(0.071)	(0.059)	(0.037)
Constant	-0.702***	-0.655***	-1.469***
	(0.052)	(0.056)	(0.028)
Observations	2,895	3,419	13,601
Log likelihood	-1,727.732	-2,164.099	-5,795.040
Note:		*p<0.1; **p<0.05; ***p<0.01	

To describe data on entry probabilities, this table shows results from a probit regression of the probability of entering each phase of development given observed drug characteristics and conditional on successful completion of the previous phase.

Overall, oral drugs, IV drugs, and drugs to treat infection have higher probabilities of entry in all phases.

Components of CBO's Structural Model and Data Sources

CBO's estimation uses the maximum likelihood that drugs enter phases I, II, and III. For each drug, CBO observes predicted lifetime revenues, predicted capitalized costs for each phase of development, and the predicted success rate for each phase of development. For a subset of drugs in the data, CBO also observes estimated actual lifetime revenues or capitalized costs.

Model component	Data source
Revenues and revenue coefficients	SSR Health
Capitalized costs and cost coefficients for each phase of development	Wouters, McKee, and Luyten (2020)
Success rate coefficients for each phase of development	Cortellis database linked with ClinicalTrials.gov
Entry decisions for each phase of development	Cortellis database linked with ClinicalTrials.gov

In each data set, CBO observes information about the drug candidate's route of administration, its therapeutic class, and whether the drug is a biologic.

Structural Estimation

Step 1. Estimate coefficients to derive revenues, capitalized costs for each phase of development, and success rates for each phase of development. Calculate revenues and capitalized costs for linked data.

Step 2. Use maximum likelihood to estimate the parameters of the distribution of $\mathrm{F} = \mathcal{N}(\mu, \Sigma)$, where $\mu = (\mu_{\mathrm{R}}\ \mu_3\ \mu_2\ \mu_1)$ and

$$\Sigma = \begin{bmatrix} \sigma_{\mathrm{R}}^2 & \rho_{\mathrm{R}3}\sigma_{\mathrm{R}}\sigma_3 & \rho_{\mathrm{R}2}\sigma_{\mathrm{R}}\sigma_2 & \rho_{\mathrm{R}1}\sigma_{\mathrm{R}}\sigma_1 \\ \rho_{\mathrm{R}3}\sigma_{\mathrm{R}}\sigma_3 & \sigma_3^2 & \rho_{32}\sigma_2\sigma_3 & \rho_{31}\sigma_1\sigma_3 \\ \rho_{\mathrm{R}2}\sigma_{\mathrm{R}}\sigma_2 & \rho_{32}\sigma_2\sigma_3 & \sigma_2^2 & \rho_{21}\sigma_1\sigma_2 \\ \rho_{\mathrm{R}1}\sigma_{\mathrm{R}}\sigma_1 & \rho_{31}\sigma_1\sigma_3 & \rho_{21}\sigma_1\sigma_1 & \sigma_1^2 \end{bmatrix}$$

assuming latent log revenues and log capitalized costs are distributed F. Parameters are identified from the observed entry to each phase of development and the observed revenues and capitalized costs and from assuming the dynamic decision-making model presented above.

μ are the means of the revenues (subscript $_{\mathrm{R}}$) and phase-specific costs (subscripts $_{3,\,2,\,1}$). Σ is the variance-covariance matrix of revenues and phase-specific costs. The diagonal shows the variances of the revenues and phase-specific costs. Other cells show the covariances between revenues and phase-specific costs.

Next Steps and Areas Where Additional Research Would Be Helpful

Next Steps: Improvements to the Model and Simulations

- Incorporate information about expected competition within therapeutic classes to improve CBO's estimates of lifetime revenues.
- Compare the output of the new model with previously released results of the former model.
- Run simulations to estimate the effects of additional policies that would affect drug development (for example, policies that would increase funding for the National Institutes of Health, expand the use of accelerated approval, or allow researchers to run less expensive clinical trials).

Areas Where Additional Research Would Be Helpful: Indication-Specific Approvals

CBO's model estimates the effects of policy proposals on the number of new drugs coming to market. The updated model will allow the agency to estimate the number of such drugs by characteristic, including therapeutic class.

CBO's model does not provide information about the approval of drugs for new indications. Additional research on indication-specific approvals could allow the agency to consider whether and how to expand the model.

- How do various policies affect firms' decisions about which indications to target for the Food and Drug Administration's approval and in which order?

- How do changes in firms' strategies regarding which indications to pursue and when affect revenues and costs and, in turn, the number of new drugs?

Areas Where Additional Research Would Be Helpful: How Changes in Drug Development Affect Health

CBO's model does not provide information about how changes in the number of new drugs affect health outcomes. Additional research in that area could allow the agency to consider whether and how to expand its model.

- How do policies that change the number of new drugs coming to market change the efficacy of the mix of drugs that are available?

- How do policies that change the number of new drugs affect health outcomes? Which patients are most affected by changes in the number of new drugs coming to market?