Commercializing biomedical research through securitization techniques

SUPPLEMENTARY INFORMATION

Jose-Maria Fernandez[1], Roger M. Stein[1,2], Andrew W. Lo[1,3,4]

[1]*MIT Sloan School of Management and Laboratory for Financial Engineering, 100 Main Street, E62–618, Cambridge, MA 02142, USA*

[2]*Moody's Corporation, 7 World Trade Center, New York, NY 10007, USA*

[3]*MIT CSAIL and EECS, 32 Vassar Street, 32–G608, Cambridge, MA 02139, USA*

[4]*AlphaSimplex Group, LLC, One Cambridge Center, Cambridge, MA 02142, USA*

Supplementary Analytics

Consider a single Bernoulli trial I_i with probability p of success ($I_i = 1$) and probability $1-p$ of failure ($I_i = 0$). Then the probability of at least one success in n independently and identically distributed (IID) trials is:

$$\Pr\left(\sum_{i=1}^{n} I_i \geq 1\right) \;=\; 1 - \Pr\left(\sum_{i=1}^{n} I_i = 0\right) \;=\; 1 - (1-p)^n \,. \tag{1}$$

For $p = 0.05$ and $n = 150$, this probability is $1 - 0.95^{150} = 0.9995$.

Assume that the drug-development process takes 10 years, and a success implies annual net income of X per year from years 11 to 20 (see Figure 1). We assume a 10-year income flow to acknowledge the fact that patents are filed long before drugs are approved, leaving much shorter periods of exclusivity than the stated 20-year life of U.S. patent. The date-10 present value Y_{10} of this stream of cashflows is given by:

$$Y_{10} \;=\; \frac{X}{r}\left(1 - \frac{1}{(1+r)^{10}}\right) \tag{2}$$

where r is the cost of capital associated with cashflows $\{X\}$. For $X = \$2$ billion and $r = 0.10$, $Y_{10} = \$12.3$ billion.

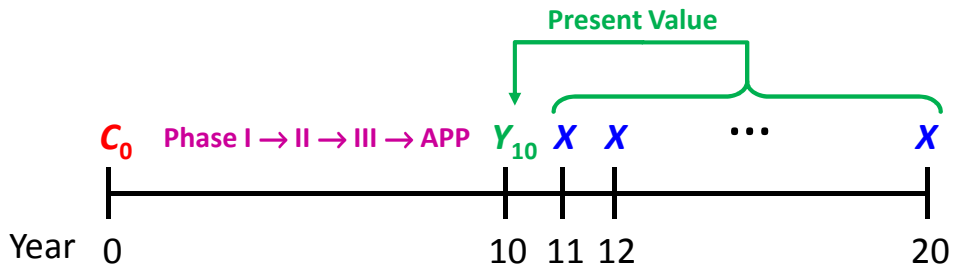

Figure 1: Timeline of cashflows for illustrative example of a typical drug development program in which out-of-pocket costs with present value C_0 at date 0 generates annual net income of X in years 11 through 20, which has a present value of Y_{10} at date 10.

If the date-0 present value of all out-of-pocket costs required to generate X is C_0, the total investment return R_s between date 0 and date 10 for a successful outcome is:

$$R_s \;=\; \frac{Y_{10}}{C_0} \;-\; 1 \;=\; \frac{X}{rC_0}\left(1-\frac{1}{(1+r)^{10}}\right) \;-\; 1\,. \tag{3}$$

The annualized rate of return over this period is:

$$\text{Annualized Rate of Return} \;=\; (1+R_s)^{1/10} \;-\; 1\,. \tag{4}$$

If $C_0 = \$200$ million, the annualized rate of return of a success is 51%.

The rate of return R for this project can then be expressed as the following Bernoulli random variable:

$$R \;=\; \begin{cases} R_s & \text{with probability } p \\ -1 & \text{with probability } 1-p \end{cases} \tag{5}$$

where R_s is defined in (3). The mean and standard deviation of R then follow directly:

$$\mathrm{E}[R] \;=\; p(1+R_s) \;-\; 1 \tag{6}$$

$$\mathrm{SD}[R] \;=\; \sqrt{\mathrm{Var}[R]} \;=\; (1+R_s)\sqrt{p(1-p)}\,. \tag{7}$$

The annualized expected return and standard deviation of return during dates 0 to 10 are then:

$$\text{Annualized Expected Return} \;=\; (1+\mathrm{E}[R])^{1/10} \;-\; 1 \tag{8}$$

$$\text{Annualized Standard Deviation of Return} \;=\; \mathrm{SD}[R]/\sqrt{10}\,. \tag{9}$$

Given the parameters assumed so far, the annual expected return and standard deviation are 11.9% and 423.5%, respectively.

For a portfolio of n identical projects that are statistically independent, the return R_p and its first two moments are:

$$R_p \;=\; \frac{Y_{10}\sum_{i=1}^{n} I_i}{nC_0} \;-\; 1 \;=\; \frac{(1+R_s)}{n}\sum_{i=1}^{n} I_i \;-\; 1 \tag{10}$$

$$\mathrm{E}[R_p] \;=\; p(1+R_s) \;-\; 1 \tag{11}$$

$$\mathrm{SD}[R_p] \;=\; \sqrt{\frac{(1+R_s)^2}{n}p(1-p)} \;=\; (1+R_s)\sqrt{\frac{p(1-p)}{n}}\,. \tag{12}$$

As (11)–(12) demonstrate, the expected return of the portfolio is invariant to the number of programs, but the risk of the portfolio (as measured by standard deviation) declines as the number of projects increases at a rate of $1/\sqrt{n}$. The annualized values of the expected return and standard deviation are given by (8)–(9) as before. For $n=150$, the return standard deviation is 34.6%.

To compute the default probability of a debt-financed portfolio of projects with total asset return R_p, we require the probability distribution of $\sum_{i=1}^{n} I_i$, which is a binomial random variable with distribution function:

$$\Pr\left(\sum_{i=1}^{n} I_i \le k\right) \;=\; \sum_{j=0}^{k}\binom{n}{j}p^j(1-p)^{n-j}\,. \tag{13}$$

The default probability of a 10-year bond at date 0 which pays no coupons and promises to pay F upon maturity at the end of year 10 is then:

$$\Pr\left(Y_{10}\sum_{i=1}^{n} I_i < F\right) \;=\; \Pr\left(\sum_{i=1}^{n} I_i < F/Y_{10}\right) \;=\; \sum_{j=0}^{\lceil F/Y_{10}-1\rceil}\binom{n}{j}p^j(1-p)^{n-j} \tag{14}$$

where $\lceil F/Y_{10}-1\rceil$ denotes the smallest integer greater than or equal to $F/Y_{10}-1$ (note that $\lceil F/Y_{10}-1\rceil+1$ is the minimum number of successful projects needed to repay the debt F), and we assume that F satisfies the inequality $0 < F/Y_{10} \le n$. For $n=150$ and $p=0.05$, the probability of default for $F=\$24.6$ billion is simply the probability of less than 2 successes out of 150 trials, which is 0.00405 according to (14). Note that the large magnitude of Y_{10} creates discreteness in debt capacity and default probabilities that may not exist in practice. For example, if $F = \$36.9 = 3\times\$12.3$ billion, the default probability jumps to 0.0182 (the probability of less than 3 successes). More generally, the debt capacity F^* associated with a desired maximum

probability δ is given implicitly by the solution to the following:

$$\max_{F} \Pr\left(\sum_{i=1}^{n} I_i < F/Y_{10}\right) \;\leq\; \delta\,. \tag{15}$$

For expositional clarity, we have assumed that the n projects are statistically independent. In practice, even the most diverse set of translational medical programs will exhibit some pairwise dependence, reducing the diversification benefits of the portfolio and, consequently, the debt capacity F^*. We incorporate such correlation explicitly in the simulations described below.

For a non-blockbuster numerical example, i.e., one in which the revenues are below $1 billion but the success rate is higher and the out-of-pocket costs of development are lower, consider the case in which the out-of-pocket cost of developing a single drug is $100 million, its expected revenue is $500 million per year for 10 years, and the probability of success is 10%. In this case, a portfolio of 100 such programs that are statistically independent would require $10 billion, yield an expected rate of return of 11.9% and a standard deviation of 29.1%, and each successful program would generate $3.1 billion in net present value in year 10. The probability of at least 4 hits out of 100 is 99.2%, implying that up to $12.3 billion of high-credit-quality 10-year debt could be issued to finance this portfolio. Such a debt issue would imply year-0 proceeds of $8.4 billion at the February 2012 Aaa yield of 3.85%. In other words, over 80% of the required $10 billion can be financed by high-quality long-term debt in this case.

Supplementary Methods: Credit Enhancement

The risk borne by investors participating in securitization transactions can be reduced using a number of protective features called *credit enhancement* mechanisms. Here we describe two types of credit enhancement that may be used individually or in combination.

The first involves the implementation of various types of structural features such as overcollateralization (through the tranching of the capital structure into different classes of securities) to increase the collateral support available for more senior bondholders, or cashflow redirection rules and triggers that accelerate payments to the more senior bondholders when certain test ratios are breached. We discuss an example of these ratios and rules in Supplementary Methods: Simulations.

The second type of credit enhancement makes use of some form of external credit support in which a third party assumes some of the risks to bondholders (e.g., through a letter of credit or bond insurance). A particularly interesting form of external credit enhancement in our context involves various forms of governmental guarantees. There are precedents for such programs. For example, in the U.S. the government sponsored enterprises (GSEs) Fannie Mae and Freddie Mac were created to promote home ownership. The GSEs provide guarantees for the mortgages underlying the securitization transactions that participate in their programs. For a mortgage to be accepted as collateral in a pool of securitized assets, it has to conform to certain quality standards

defined by the GSE. A recent example from the biomedical domain is the Israeli Life Sciences Funds, a venture capital fund of over $200 million jointly launched in 2011 by several branches of the Israeli government and the private sector. To mitigate the risks associated with biotech R&D, the government will assume some of the downside risk.

In light of these precedents, the National Cancer Institute or the National Institute for Health could consider providing some form of guarantee to biomedical megafunds whose collateral conformed to some pre-defined scientific or medical criteria. Alternatively, a private foundation might assume this role.

Credit support from such a benefactor could serve to boost investors' interest in these securities and potentially allow the megafund to assume bigger risks in its investments (e.g., by investing in newer technologies or those with less certain outcomes or that target rarer diseases) while providing a mechanism for leveraging the capital available from the guarantor or benefactor.

Supplementary Methods: The Drug Approval Process[1]

The introduction of a new drug in the market is a highly regulated process. Countries typically have national agencies responsible for authorizing new compounds for sale. For the purposes of this study, we follow the process defined by the U.S. Federal Drug Administration (FDA). Every new pharmaceutical product must undergo a number of tests to ensure that it is safe and effective. The lifecycle of a new drug generally follows the path described below.

In the "Preclinical Phase" the company developing the drug tests the product in animal trials to produce evidence that there is reasonable cause and manageable risk to permit the compound to proceed to human studies, in accordance with FDA guidelines. Following this phase, the sponsoring company files an "Investigational New Drug" (IND) application. If the FDA approves the IND, the drug moves into "Phase I", in which the drug is tested in a small number of healthy volunteers to monitor its absorption, metabolism, and toxicity in the body to get information about its safety and dosage. If the drug is determined to be too toxic or otherwise unfit, it is withdrawn at this point.

Compounds that successfully pass Phase I move into "Phase II", where testing is done with a patient population that already has the disease targeted by the new compound. The sponsor of the trial defines a set of endpoints that exemplify the compound's desired effectiveness and compares these endpoints with the results from the trials in diseased patients.

Upon successful completion of Phase II, the drug moves into "Phase III" in which the drug is tested in a large sample of patients to try to confirm safety and efficacy in a wider number of circumstances and subjects.

Following successful completion of these trials, the sponsor may submit a "New Drug Application" (NDA) or "New Biologics Application" (BLA) to the FDA. If the NDA or BLA is approved, the drug can be legally marketed in the U.S.

Supplementary Methods: Simulations

In this section, we describe the specific assumptions and experimental design used to generate the simulated performance analysis for an oncology megafund. The goal of our simulations is not to define an optimal transaction structure or to defend a specific set of modeling assumptions. Instead, our intent is to demonstrate the feasibility of modeling a simple financial structure—using realistic economic and scientific assumptions—in which large-scale biomedical innovation yields potentially attractive investment and drug development properties. To encourage readers to experiment with the simulation, we provide the complete sourcecode in R and Matlab under an open-source license that enables researchers to use, modify, and distribute it.

Time Units and Tenor

The time unit used in our simulations is a semester (six months). Alternative time steps are possible but we chose one semester to match the semiannual coupon payments of the bonds in the fund's capital structure.

We assume that the life of the fund is 15 semesters (seven and a half years). The scheduled amortization for the bonds occurs in periods 5–8 for the senior bond and 9–12 for the junior bond (we use the term "period" to refer to particular semesters). Thus, the longest-dated bond issued has a tenor (time between issuance and maturity date) of six years. In the 15th period, any remaining assets that have not been either already sold or discontinued are sold and the revenues generated accrue to the equityholders. We have assumed that it takes one year to sell a compound. Consequently, equityholders would get paid at the end of period 17 of the simulation, provided there were no previous defaults that might have shortened the life of the megafund.

Funds and simulations can be easily structured for significantly longer durations. The tenor of the fund should be related to the expected time required for the largest number of compounds to reach their full economic value. Given that Simulations A and B replicate the development of compounds from Preclinical and Phase I to Phase II, and from Phase II to market approval, respectively, we obtained reasonable results for funds of about seven and a half years' duration.

Assets

The assets in the portfolio are assumed to be new drugs being developed by biotech or pharma companies and targeted at curing some form of cancer. Under the current model design, those same companies would be responsible for developing the compounds. The megafund could act as a financing partner and a platform from which funding could be structured, subject to a set of rules that foster collaboration across projects, encourage individual and group success, and avoid moral hazard. Those rules, as well as the processes required to select and manage the assets targeted for the fund (including determining which assets to sell and in which phase of the process to sell them), will need to be defined for each new megafund, depending on its objectives and structure. Finding the right balance between the protection of investors and the development of new scientific

solutions is one of the difficult tasks regarding the implementation of this model. The creation of a blue ribbon Scientific Committee supported by financial experts is a necessary condition for the success of this new type of vehicle. The megafund structure provides an opportunity to revisit the way drug development financing decisions are currently made and to explore new corporate governance structures and organizational designs.

In our experiments, the initial portfolio of assets is composed of compounds in either the Preclinical and Phase I stages (Simulation A) or in Phase II (Simulation B). Simulation A replicates the typical venture capital investment horizon that carries compounds from Preclinical or Phase I to Phase II, when a large pharmaceutical company may acquire or license the compound for later development. Simulation B replicates the subsequent biopharmaceutical investment and development of a compound from Phase II to market approval. Practitioners note that different skill sets and funding budgets are required for each of those horizons, which motivated our decision to split the simulation in this manner. Taken together, the two simulations provide a compelling case for applying megafund financing throughout the full lifecycle of compound development.

We assume that all compounds are acquired during the first semester of the life of the fund and that no new compounds are acquired thereafter. The number of compounds in each of the simulations results from a two-step process: (1) we fix an amount of equity such that the all-equity fund and the RBO fund both have the same dollar amount of equity ($2.5 billion in Simulation A and $6 billion in Simulation B); (2) we determine the maximum number of compounds that we can expect to invest in using all the equity and debt raised in each of the funds. We acknowledge that this is a strong assumption. Deploying several billions of dollars in such a short period of time would require considerable prior due diligence work, some new form of collaboration with the current market players who have developed the expertise in making these types of investments (venture capitalists, biopharma companies, etc.), or a totally new approach to allocate capital across development drugs. While it may be more effective to deploy capital in a less abrupt fashion, i.e., acquiring new compounds throughout the life of the fund, modeling a dynamic and actively traded portfolio would create greater complexity in this simulation experiment. Therefore, for the purposes of our examples we adopt the simpler approach of acquiring the fund's collateral upfront, as is common among current securitization transactions. The composition of the portfolios in our simulation experiments are given in Table 1.

	Simulation A		Simulation B	
	Equity	RBO	Equity	RBO
Assets				
Preclinical	50	100	—	—
Phase I	50	100	—	—
Phase II	—	—	40	100
Phase III	—	—	—	—

Table 1: Composition of initial portfolios of drug compounds in the simulations.

Even though all assets are acquired at date 0, not all of the cash available at date 0 is invested immediately. A cash reserve is required to finance future clinical trials whenever a compound

transitions into a new clinical phase. The fund reserves as much cash as will be required (in expectation) to develop the compounds and to cover the interest payments on the notes for a certain number of semesters. In our simulation experiments, we reserve cash for two semesters' worth of interest payments, i.e., the bonds do not amortize in the first year.

Simulating Portfolio Dynamics

The development of a new drug is a complex process that depends on various scientific and economic factors. Our simulation is based on the assumption that every compound can transition along a series of different predefined states of the approval process: Preclinical, Phase I, Phase II, Phase III, NDA or BLA, Market Approval, and Discontinuation. In our model Discontinuation and Approval are absorbing states, i.e., a drug that is discontinued or approved can no longer transition into any other state. The assumption that drugs cannot be discontinued after being approved is explained by the fact that our megafund would sell all approved drugs to biopharmaceutical companies to be marketed immediately after being approved for their first indication. Also, in our data we observe a small probability that in some rare cases a compound may skip a phase—for example, conducting trials in Phase I and II simultaneously and transitioning directly to Phase III—so our simulation captures this behavior as well.

We simulate the evolution of compounds through the approval process as following a Markov process, which is a common modeling tool applied to systems that transition from state to state over time. This approach has been widely used in finance to represent various forms of credit risk,[2] and we apply it in our context to model the drug-transition dynamics from one phase to the next.

For each state in the approval process, we estimate a vector of transition probabilities based on the analysis of data from the last 20 years of oncology drug development programs. The vector is made up of elements, p_{ij}, each of which represents the probability that a compound transitions from state i to state j in the next time step (one semester in our simulations).

At every time step and for each compound that is still in the portfolio, i.e., the compound has not been discontinued or sold, we generate a random number u from a uniform distribution. The value of u is then compared to the vector of probabilities p_{ij} to determine whether a transition occurs in the next period and if so, to which state. Conceptually, this works similarly to spinning a roulette wheel where the slots represent the different states in the approval process and the size of each slot is proportional to the probability of being in that state one period later.

In the event that a transitioned compound ends up in either an Approval or Discontinuation slot, the compound is sold or dropped from the portfolio, respectively. Otherwise, it continues in the process, but now uses a new probability vector for the new state to which the compound has transitioned.

We implement simple rules to determine which compounds to sell during the life of the fund. In particular, we assume that all drugs that are approved are sold to biopharmaceutical companies who will ultimately be marketing and distributing them. In addition, compounds can be sold prior to approval to meet the interest, principal, or management fee payments. The values and other features of the sale process are described below.

Transition Probabilities

The transition probabilities were calculated in two ways. For compounds in Phase I or later, we used a research database that we constructed for this study. For Preclinical compounds, we refer to existing literature for the relevant probabilities and adjust them for the periodicity of our simulation.[3]

For the compounds in the clinical development phases of our simulation, the transition probabilities were calibrated using two sources of historical data: the DEVELOPMENT optimizer™ database provided by Deloitte Recap LLC and a dataset provided by the Center for the Study of Drug Development (CSDD), Tufts University School of Medicine. Recap's DEVELOPMENT optimizer™ database is built on curated clinical and regulatory histories for approximately 1,450 compounds entered into human clinical development in 2,467 distinct indications by a select group of more than 240 benchmarked biotechnology companies, i.e., the constituents of the Recap Bio-Portfolio Index™, since 1988. The histories are documented and updated daily using multiple primary, public sources of information, including but not limited to: U.S. Securities and Exchange Commission filings, U.S. and E.U. pharmaceutical regulatory documents captured and analyzed from the Food and Drug Administration (FDA) and the European Medicines Agency websites, peer-reviewed journal articles and scientific abstracts, government databases such as http://clinicaltrials.gov, and corporate press releases and investor presentations. The CSDD data were compiled from publicly available information reported by companies involved in the development of cancer drugs. The compounds targeted consisted of new molecular entities developed primarily for an anti-cancer indication for which an IND application was filed with the FDA and that entered clinical trials between January 1990 and the start of 2011. The compounds in the database were developed by biotechnology or pharmaceutical companies and were either therapeutic compounds or vaccines.

We merged the Recap and CSDD databases to yield a combined database of over 2,000 compounds. After removing duplicates and compounds for which there was not enough information about their start or transition dates, or that did not conform to the criteria defined in the paragraph above (e.g., compounds only approved for marketing outside of the U.S. or that were reformulations of existing drugs), we arrived at a final set of 733 compounds. The summary statistics for this final database are contained in Table 2.

Using these data, we calculated the transition probabilities by first estimating a continuous-time generator matrix and then converting this to transition matrices in a discrete-time setting where the time unit is the semester.[4]

For compounds in the Preclinical phase, data is more difficult to collect, in part because drug development companies have little incentive to provide information about their research and sometimes unsuccessful programs. To estimate transition probabilities for Preclinical compounds, we used statistics reported in Paul et al. (2010) and also adopted the definition of the Preclinical period used in that paper.[3] We then scaled these long-run probabilities to arrive at the probabilities for a single semester. In doing so, we assume that compounds in this phase could only transition from Preclinical to either Discontinuation or Phase I, and that the mean time in the Preclinical phase was one year, as reported in Paul et al. (2010).

Stage	Total	in %
Approved:	38	5%
Discontinued (NDA)	2	0%
Discontinued (Phase I)	174	24%
Discontinued (Phase II)	171	23%
Discontinued (Phase III)	30	4%
Still in process as of end compilation period:		
In NDA	4	1%
In Phase I	17	2%
In Phase II	221	30%
In Phase III	76	10%
Total	733	100%

Table 2: Composition of the final database of 733 oncology compounds in various clinical phases (percentages do not sum to 100% due to rounding).

The resulting transition matrix estimate $\mathcal{P}$ is given in Table 3 and the mean long term transition probabilities (the limiting probabilities) and mean times to transition resulting from this probability matrix over the period of time covered by the study are presented in Table 4. These phase transition probabilities are comparable to those reported elsewhere in the literature for similar compounds, as shown in Table 5.[5–8]

$$\mathcal{P} = \begin{array}{c} \\ \text{Preclinical}_t \\ \text{Phase I}_t \\ \text{Phase II}_t \\ \text{Phase III}_t \\ \text{NDA}_t \\ \text{Approved}_t \\ \text{Withdrawn}_t \end{array} \begin{array}{c} \begin{array}{ccccccc} \text{Preclinical}_{t+1} & \text{Phase I}_{t+1} & \text{Phase II}_{t+1} & \text{Phase III}_{t+1} & \text{NDA}_{t+1} & \text{Approved}_{t+1} & \text{Withdrawn}_{t+1} \end{array} \\ \begin{pmatrix} 50.0 & 34.5 & 0.0 & 0.0 & 0.0 & 0.0 & 15.5 \\ 0.0 & 80.8 & 13.3 & 0.5 & 0.0 & 0.0 & 5.3 \\ 0.0 & 0.0 & 84.5 & 6.7 & 0.3 & 0.1 & 8.5 \\ 0.0 & 0.0 & 0.0 & 84.8 & 6.8 & 2.1 & 6.3 \\ 0.0 & 0.0 & 0.0 & 0.0 & 56.7 & 41.2 & 2.2 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 100.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 100.0 \end{pmatrix} \end{array}$$

Table 3: Estimated transition matrix used in all simulations (in percent). Time subscript t indicates current six-month simulation period and $t+1$ indicates the following six-month simulation period. Note: entries in each row do not necessarily sum to 1.00 because of rounding.

Asset Valuations

In this section, we discuss the distributional model we use for simulating the values of drug compounds at the time of sale. An important feature of our model is the presence of correlation among

	Preclinical to Phase I	Phase I to Phase II	Phase II to Phase III	Phase III to NDA	NDA to Approval
P (Transition)	69.0%	72.4%	45.2%	58.6%	95.2%
Avg (months in phase)	12.0	31.2	38.6	39.6	13.8

Table 4: Average transition probabilities and time per development phase.

Source	Time Period	Number of Compounds	Preclinical to Phase I	Phase I to Phase II	Phase II to Phase III	Phase III to NDA	NDA to Approved
Megafund*	1990–2010	733	69.0%	72.4%	45.2%	58.6%	95.2%
Natanson*	1988–May 2010	164	—	72.6%	40.3%	66.7%	90.6%
Reichert et al.*	1990–2006	920	—	78.0%	43.0%	52.0%	89.0%
Walker et al.*	1995–2007	974	—	77.0%	44.0%	52.0%	—
Dimasi et al.	1993–2002	838	—	76.8%	59.4%	57.1%	—
Paul et al.	15 years	—	69.0%	54.0%	34.0%	70.0%	91.0%

*These probabilities are calculated only for cancer related compounds.

Table 5: Comparison of cancer compound transition probability by development phase.

the valuations of different compounds, which can be observed empirically to some degree and is often noted by venture capitalists and other experts in this domain. We begin by assuming that the market values of drug compounds are approximately lognormally distributed, which implies a larger number of moderately low valuations interspersed with a smaller number of large "blockbuster" valuations. In addition, we induce correlation among the valuations of compounds in the collateral portfolio. Although correlation does not affect the mean return on the overall portfolio, because of the overlay of tranching and the impact of the waterfall rules, the mean return and variability of returns on *specific tranches* will be affected by contemporaneous clusters of particularly high or particularly low valuations.

To induce correlation among the sale values in our model, we begin by generating correlated standard normal random variables which we then rescale and exponentiate such that the resulting (correlated) lognormal random variables have means and standard deviations that are consistent with the valuation assumptions described in Supplementary Methods: Simulations (Calibration of Valuation Parameters). This mechanism may be thought of as specifying a stochastic process for the value of a compound in which there is a common unobservable factor.

Next, we specify that the value Z_{ij} of the standard normal draw for the ith entity in the jth economic state of the world is the sum of two components: (1) a common (systematic) component that affects all valuations in the portfolio in economic state of the world j; and (2) a compound-specific (idiosyncratic) component that only affects compound i in economic state of the world j. More formally:

$$Z_{ij} \;=\; \beta_i^S S_j \;+\; \epsilon_{ij} \qquad , \qquad S_j \;\sim\; \mathcal{N}(0,1) \;,\; \epsilon_{ij} \;\sim\; \mathcal{N}(0, 1-(\beta_i^S)^2) \tag{16}$$

and β_i^S describes the magnitude of the impact of the systematic factor S_j on Z_{ij}. We assume that all error terms and cross terms are mutually statistically independent.

Under these assumptions, for compounds with a common value $\beta_i^S = \beta_S$, and recalling that $Z_{ij} \sim \mathcal{N}(0,1)$, it can be shown that

$$\beta_s \;=\; \sqrt{\rho} \qquad , \qquad \epsilon_{ij} \;=\; \sqrt{1-\rho}\,\eta_{ij} \tag{17}$$

where $\eta_{ij} \sim \mathcal{N}(0,1)$ is IID over i and j, and ρ is the correlation between the Z_{ij} (note: ρ must be greater than or equal to 0 and less than or equal to 1). This specification is similar to common one-factor credit models with correlated defaults (however, in our case the transition probability is not driven directly by the asset value—as might be the case in models of correlated default—since the transition probability is determined primarily by the drug approval process).[9] In cases in which there is a common correlation among all compounds in the portfolio ($\beta_i^S = \beta_S$ for all i), the value of the ith compound in the jth simulation path is given as

$$X_{ij} \;=\; \exp\left(\mu + \left(\sqrt{\rho}S_j + \sqrt{1-\rho}\epsilon_{ij}\right)\sigma\right) \;=\; \exp\left(\mu + Z_{ij}\sigma\right) \tag{18}$$

where

$$\begin{aligned} \mu &= \ln(m_v) - \tfrac{1}{2}\ln\left(1+\tfrac{s_v^2}{m_v^2}\right) &&\equiv \text{ the estimated mean of log-returns} \\ \sigma &= \left[\ln\left(1+\left(\tfrac{s_v^2}{m_v^2}\right)\right)\right]^{1/2} &&\equiv \text{ the estimated SD of log-returns} \end{aligned} \tag{19}$$

and m_v and s_v^2 are the estimated mean and variance, respectively, of the observed valuation data (or are derived from other qualitative approaches).

To avoid very large (and potentially unrealistic) simulated values for compounds at the time of sale, we also impose a maximum (M_i) on the value of the compound such that the final value is given as

$$V_{ij} \;=\; \min(X_{ij}, M_i)\;. \tag{20}$$

The introduction of an upper bound affects the mean of the distribution. We adjusted the values of μ to accommodate the capping that occurs as a result of the imposition of the upper bound. This adjustment ensures that the mean of the distribution is consistent with our data. The values for μ, σ, and M_i are presented in Table 6 for each compound phase.

In our simulations, drug compounds are sold infrequently and tend not to cluster in time except at the end of the transaction. Thus, simulations in which the (unobservable) systematic component is updated each period result in relatively little correlation among prices, even though our simulation horizons are short (i.e., on the order of 5–7 years). Furthermore, because the current structure of the portfolio is static, i.e., the portfolio of compounds is purchased at the beginning

of the fund transaction and then winds down, it is natural to think of a common factor affecting the whole portfolio. Such a factor could be general economic conditions, regulatory shifts, or sweeping technological advances.

Accordingly, we assume that the valuations of compounds are lognormally distributed and governed by the dynamics in (16), and that the systematic shock occurs once at the beginning of each simulation path, such that a particular string of sales within that path will *all* be influenced by the shock. This induces correlation among all valuations in the portfolio, rather than just between those in which sales occur in a single period.

To implement this approach, in each path we simulate a single value for S_j (in the jth path) and then, for each compound being sold, simulate ϵ_{ij} and calculate a valuation for compound i (V_{ij}) as in (20). Importantly, although S_j is drawn only once per trajectory, the compounds in the portfolio themselves evolve (e.g., transition, get funded, etc.) in each time step of the simulation.

Calibration of Valuation Parameters

A critical set of assumptions in our simulations involves the valuations of individual drug assets at a given stage of development. Such valuations are difficult to estimate due to significant heterogeneity across the assets with respect to a compound's scientific merits, its commercial potential, the expertise of the managers in charge of its development, etc. Furthermore, many oncology assets have traditionally been privately held, or developed as part of a larger suite of products, and thus accurate data on individual valuations are not readily available.

To estimate the mean and variance of compound valuations for our model, we used data from Bloomberg to build a dataset of initial public offerings (IPOs) since 2000, and market valuations (as of the end of the first quarter of 2011) of publicly listed companies in the U.S. focused on developing and marketing cancer compounds and which had a market capitalization of at least \$5M. For each company for which we could gather enough information, the value per approved compound was calculated by dividing the market capitalization of the company by the number of approved compounds it owned. We realize that this method may overestimate the value of compounds given that this calculation assigns a value of zero to earlier stage compounds and that the company may hold other assets not considered. Alternate methods to estimate the value of compounds such as discounted future cashflows could be applied to future developments of this model. For our simulations we decided to approximate the value of a compound using observed market valuations. The resulting mean value for a marketed compound is \$1.87 billion and its standard deviation is \$2.24 billion.

The valuations corresponding to the compounds in earlier development phases were calculated using a binomial-tree valuation model in which the value of each compound is estimated by taking into account the probabilities of success and failure per phase, the expected values in each case, and the time required to move from one phase to the next.[10] The inputs for the transition probabilities and times in each phase were derived from our transition matrix. The discount rates used to calculate the discounted values per phase were 15% for the Market to NDA phase, 25% for NDA to Phase III, and 30% for the earlier phases to reflect the higher risk of early-stage projects. In addition, upper bounds for valuations were imposed to prevent the model from generating un-

reasonably large values. These upper bounds were chosen qualitatively based on the empirical distribution of values (see Supplementary Methods: Simulations (Asset Valuations)).

To compute the standard deviation of each of the values per phase and per compound, we used data from Bloomberg. First we calculated the standard deviation of the value of marketed compounds (the $2.24 billion cited above). Next we calculated the ratio (1.19) of the standard deviation to the mean value of marketed drugs (the $1.87 billion cited above) in our Bloomberg database. Finally, we applied this ratio to the mean values per compound and per phase to estimate each of the standard deviations corresponding to each of the phases.

The mean, standard deviation, and upper bounds were used to fit the lognormal distributions from which the value of each compound is drawn in our simulations. Table 6 shows the values of the estimated original mean and upper bounds as well as the cap-adjusted μ and σ used to estimate the value of the drugs.

			Cap Adjusted	
	Original	Cap	μ	σ
Preclinical	16	100	2.36	0.939
Phase I	30	250	2.96	0.939
Phase II	82	500	4.00	0.939
Phase III	425	1000	5.80	0.939
NDA	1515	2500	7.35	0.939
Approved	1870	5000	7.24	0.939

Table 6: Parameters for valuation functions, where μ and σ correspond to the mean and standard deviation of the lognormal distribution from which valuations are randomly drawn.

Finally, as a proxy for the (normal) correlations, we use the mean pairwise correlation of the equity returns on small biopharmaceutical firms, also calculated using Bloomberg data. These are related somewhat to the valuations of individual compounds since it is often the case that small firms have only a single drug that they are either researching or producing. Thus the price of the firm may be related to the value of these compounds.

In our sample, the equity correlations we estimated were on the order of 20%, which is the value we use in the simulations. This value is within the range of observed correlations among public firms in typical credit-portfolio models, and is consistent with empirical estimates of publicly traded oncology biotech firms. We expect correlations associated with very small private firms to be lower than for those of public firms, and correlations among the valuations of single compounds in our simulation to be lower still.

Note also that the correlation in returns does not translate directly into an equivalent level of correlation for realized valuations, due to the calculation of the former on returns (in normal space) and the latter on levels (in lognormal space). In general we observe that valuation levels are correlated to a lower degree than are the equity returns (i.e., the correlation among valuations is lower than 0.2), which is consistent with our expectations.

Investment Structure and Development Costs

The investment structure we assume is based on the licensing framework commonly used in the biopharmaceutical industry. In our model, during the course of the drug's development, both upfront and periodic payments are made by the megafund to finance additional research and to compensate the developers for successful completion of key milestones (such as the completion of a phase). In addition, the megafund finances all clinical trial costs. In exchange for this funding, the megafund is granted 85% of the economic value of the compound when it is sold. The remaining 15% is assumed to be retained by the founder and management team developing the compound. The structure of the payments made by the megafund is detailed below.

Phase	Upfront Payment	Milestone	Development Cost
Preclinical	2.5	1.3	Random*
Phase I	7.5	3.8	Random*
Phase II	20.0	10.0	Random*
Phase III	75.0	37.6	Random*

*See Supplementary Methods: Simulations (Drug Development Costs).

Table 7: Investment costs (in $millions).

In practice, upfront and milestone payments for a specific compound are derived through negotiations based on the novel features and properties of the compound, its expected value, the amount of investment required to carry the compound to the next phase(s), and the negotiating power of the parties. It is therefore difficult to define what the standard terms of any particular deal might be. For our model, in addition to the commitment to fund the development of the drugs, we estimated the upfront payments to be 40% of the expected development costs per phase and the milestones to be 50% of the upfront payments.

We confirmed the plausibility of these investment parameters through conversations with experts, and by using data from public presentations made by practitioners[11] and the Recap DEAL builder™ tool. However, recent trends seem to favor smaller upfront payments and larger milestone payments. Future research may confirm this point, which would necessitate changes in the simulation parameters.

Drug Development Costs

We assume that development costs per phase and per compound follow a lognormal distribution with parameters based on previous results reported in the literature.[3,12] Some authors have since argued that the costs reported in these studies may be overstated,[13] but we adopt these figures to be conservative.

For compounds in preclinical and clinical phases, Paul et al. (2010) provides estimates of the cost of development at each clinical stage based on industry benchmarking data provided by the

Pharmaceutical Benchmarking Forum along with fifteen years of project level data from Lilly's R&D portfolio. Dimasi et al. (2003) bases results on survey data collected from the commercial sponsors of 68 randomly selected approved compounds in the CSDD database, representing multiple therapeutic areas including, but not limited to, oncology.

To estimate the mean cost per phase we have chosen to use the results of Paul et al. (2010) because they are more timely and they result in more conservative expected costs. However, we omitted the submission and launch costs proposed ($40 million) which appear to include launch preparation costs that we expect to be borne by the biopharma companies acquiring these compounds post-approval (recall that under our current assumptions, once drugs are approved for their first indication, they are sold to an industry incumbent for marketing and distribution).

We further increase our cost estimates in two ways. First, since the statistics in Paul et al. (2010) are calculated in 2008 dollars, we used the U.S. GDP deflator index to inflate the numbers to 2011 dollars. Separately, we adjust for the additional cost of developing cancer compounds relative to other types of therapies. Adams and Brantner (2006)[14] analyze the capitalized cost of new drug development by indication and show that the cost of developing an oncology product is 20% higher than the sample mean (across all compounds in their dataset). Accordingly, for our analysis, we adjusted the costs upward by a factor of 1.2 to reflect the higher than average cost of oncology development. Both of these adjustments yield higher costs and, therefore, more conservative profits in our simulations.

Using our mean estimates, the resulting mean out-of-pocket costs invested per compound from Preclinical to the end of Phase III is $263 million.

Paul et al. (2010) do not provide an estimate of the standard deviation of the costs per phase. For our experiments, we assumed that the variability of development costs for oncology compounds is related to that presented in Dimasi et al. (2003). We then calculated the ratio of the standard deviation to the mean cost reported in this paper (ranging from 0.70 to 0.94) and applied that ratio to the adjusted costs per phase obtained as previously explained.

In addition, we imposed a maximum cost in each phase to cap the expenses incurred per compound and per phase. The sum of the cap costs assumed per phase yields a total maximum out-of-pocket cost per compound of $690 million, which is quite conservative compared to figures contained in the literature. The resulting adjusted mean costs per phase and corresponding standard deviations are shown in Table 8.

These figures can be used to estimate the parameters μ and σ of the lognormal distribution that we use to simulate development costs, which are drawn randomly for each compound in each phase. We adjusted the values of μ to accommodate the capping (the maximum cost per phase) to ensure that the mean of the distribution would remain consistent with the observed data. The resulting parameters used are given in Table 9.

The out-of-pocket costs per approved compound may be estimated according to Paul et al. (2010) by calculating the number of drugs needed to obtain a single approval. Under our assumptions, eight Preclinical projects are needed to bring one new drug to market on average. Starting with eight compounds in the Preclinical phase, if we multiply the expected number of compounds that transition to each phase by the expected cost per phase we get an estimate of the out-of-pocket

	Mean cost Paul et al. (2010)	Mean adjusted for oncology factor (in USD2011)	Dimasi et al. (2003) SD/Mean ratio	SD per phase	Max cost per phase
Preclinical	5	6	0.92	6	20
Phase I	15	19	0.84	16	50
Phase II	40	50	0.94	47	120
Phase III	150	188	0.70	132	500
Total		263			690

Table 8: Asset development out-of-pocket costs.

	Adjusted Mean Cost	Cap Adjusted μ	σ
Preclinical	6	1.53	0.79
Phase I	19	2.72	0.73
Phase II	50	3.65	0.79
Phase III	188	5.06	0.63
Total	263		

Table 9: Parameters μ and σ of the lognormal distribution used to simulate development costs.

cost to develop a new compound of $693 million. Following Paul et al. (2010) and Dimasi et al. (2003), we capitalize these costs over time to account for the cost borne by investors to finance the development of drugs. The resulting mean total capitalized cost per approved drug using a discount rate of 10% is $1.2 billion.

Cost	Megafund Simulation	Paul et al. (2010)	Dimasi et al. (2003)	Dimasi (2007)	Adams (2006)
Out-of-pocket cost	693	654	403	672 / 559	—
Capitalized	1,220	1,104	802	1,318 / 1,241	868

Table 10: Comparison of development costs of a single approved drug (from Preclinical to Approval).

Capital Structure and Cashflow Waterfall

In our experiments, we assume a very simple capital structure and cashflow waterfall. Our capital structure has three tranches: a senior bond, a junior bond, and an equity tranche. A more sophisticated implementation would almost certainly take advantage of a more efficient capital structure and more involved waterfall rules.

The bonds receive semiannual coupons and are amortized in equal installments over various periods of time as presented in Table 11. The senior bonds have a maturity of 4 years and their owners receive coupon and redemption payments ahead of the junior and equity-tranche holders. The junior bonds have a maturity of 6 years and they are paid back before any cashflows accrue to the equityholders.

The securities are assumed to be quite basic (fixed-rate amortizing bonds and common equity), but a more sophisticated model might make use of less standard securities to better match investors' preferences.

	Coupon (annual)	Amortization Schedule (start and end semester)
Senior Bond	0.05	Periods 5 to 8
Junior Bond	0.08	Periods 9 to 12
Equity	—	Period 15

Table 11: Capital structure parameters used in simulations.

In addition to overcollateralization (which involves holding more collateral than the par value of the tranche), the structure includes an interest coverage ratio test (IC test) designed to protect bondholders. The ratio is calculated as follows:

- The numerator of the IC ratio is equal to the cash reserve available plus the future expected cash inflows from the sale of compounds already in process minus the management fee (50 basis points per year) minus the interest and principal redemptions due in the current period.
- The denominator of the IC ratio is the required payments for the next k periods of management fees plus interest and principal (k is assumed to be 2 for Simulations A and B).

If the ratio falls below the target IC level, compounds are sold to bring the IC ratio back into compliance. The sale of assets helps ensure that the Special Purpose Vehicle (SPV) will have enough funds to pay the servicing, interest, and principal payments.

Throughout the life of the biomedical megafund, waterfall rules guide the allocation of funds. The waterfall is implemented as follows:

- At the start of each period (semester) all proceeds from any consummated compound sales are added to the current cash balance.
- Also at the start of each period, each compound is tested to see if it has transitioned to a new state. Any compounds that have transitioned into the Approved state (or to the targeted phase in the megafund, i.e., Phase II in Simulation A) are sold and the cashflow from the sale is deferred until the end of the sales cycle (we assume it takes 2 semesters to organize and execute a compound sale).
- If there sufficient cash in the cash account, payments are made in the following order:
  - The megafund management fee is paid.
  - Interest on senior bonds is paid.
  - Scheduled principal payments on senior bonds are paid.
  - Interest on junior bonds is paid.
  - Scheduled principal payments on junior bonds are paid.
- If there is not enough money to meet these obligations, some or all of the bonds are in default and the assets are liquidated. In the event of liquidation, the cash generated by the monetization of available assets flows first to the most senior bondholders followed by the junior bondholders, and any residual amount goes to the equityholders.
- If the megafund is not in default, the IC test is performed. If the IC test is failed, the cash shortfall is calculated and compounds are sold to meet the shortfall and ensure compliance with the IC test.
- From the remaining cash, a portion is reserved to make the servicing, interest, and principal payments over the subsequent k periods of time (k is assumed to be 2 in the simulations).
- If any surplus cash remains, it is used to finance the clinical trials of any compounds that have transitioned but have not yet received funding for their new phase, starting with those compounds that are farthest along in the approval process.

- After all of the above payments are made, cash at the end of the period is calculated.
- If there are no bonds left outstanding, the portfolio is liquidated and all remaining proceeds, net of administrative fees, accrue to the equityholders.

Supplementary Discussion

In this section we review some of the practical challenges of megafund financing and potential solutions. One set of issues involves our choice of simulation parameters: in some cases, they are conservative, in other cases, they are aggressive. On the conservative side, our assumed 8% and 5% coupon rates for junior- and senior-tranche research-backed obligations are higher than those required by today's investors; we assumed a probability-adjusted cost of developing a compound of over $1.2 billion; and we ignored potentially significant synergies and cost savings likely to accrue to a large entity involved with multiple anti-cancer-therapy teams. On the ambitious side, we assumed that compounds that are not discontinued can be sold within one year at some random price drawn from a lognormal distribution, and that a very large number of compounds can be developed simultaneously and efficiently by multiple teams working in parallel. To allow others to evaluate the importance of these concerns by conducting new simulations with their own choice of parameters, we have placed our simulation software in the public domain with an open-source license to run, modify, and distribute the code.

We have also made assumptions regarding the capacity of translational medical research that are harder to test via new simulations. In particular, we have implicitly assumed that there is a sufficient supply of anti-cancer compounds to meet the demands of our megafund; that several billion dollars of capital can be deployed in high-quality research programs over a short startup period; and that there is no shortage of talented researchers, engineers, and entrepreneurs who will staff the various teams needed to develop these compounds. These assumptions are partly motivated by informal discussions with numerous biomedical researchers who seem to have more innovative ideas than they have funding, and who observe that the availability of funding sometimes seems inversely related to the innovativeness of their proposed research. These assumptions are also motivated by reports of 20-year inventories of oncology compounds waiting to be investigated,[15] an increasing number of academic biotech spin-outs,[16], and the increasing number of science and engineering doctorates awarded over the past decade (with the biggest increases coming from the medical and life sciences[17]). Finally, our optimism regarding capacity is motivated by recent theoretical and empirical research in financial economics documenting the positive impact that increased investment activity has on innovation by stimulating an increase in the supply of innovators and truly novel ideas.[18–20] In one such study,[19] the authors conclude that "...the flood of capital in hot markets also plays a *causal* role in shifting investments to more novel startups—by lowering the cost of experimentation for early stage investors and allowing them to make riskier, more novel, investments."

Of course, one of the most speculative assumptions underlying our simulations is that historical drug-development data can be used to calibrate the parameters of our simulation. New trends and nonstationarities in the stochastic process of biomedical R&D may reduce the accu-

racy of such extrapolations. For example, the fantasy of personalized medicine is fast becoming a reality through advances in pharmacogenomics and the identification of genetic and molecular biomarkers for various types of cancer.[21] This recent innovation has had a dramatic impact on the biopharma industry, creating smaller and less correlated biotech niches but also inducing greater correlation among big pharma companies that are targeting the same molecular pathways.[15] More accurate targeting of drugs also affects pricing and reimbursement policies[22] which, in turn, have important consequences for biopharma revenues, business priorities, and, ultimately, research-and-development decisions. The parameters of our simulation are clearly affected by such considerations, hence these and other context-specific issues must be addressed in any live application of our approach.

We acknowledge the inherent imprecision of any extrapolation of current research agendas and business trends. However, the inability to accurately predict translational research outcomes does not imply an inability of investors to assess the financial risks of and commit capital to a diversified portfolio of such outcomes. With sufficient scale, time, and expertise, biomedical megafunds may well yield attractive investment opportunities to a much broader universe of investors than those who currently invest in the biopharma industry. Securitization addresses scale and time, but obtaining the necessary expertise requires an unprecedented level of collaboration between academia and industry, and among doctors, scientists, and engineers,[23] including financial engineers. By incorporating more specific knowledge about industry trends, transformative scientific discoveries, and potential interactions between various drug-development programs into a megafund's portfolio construction process, investment performance can be improved substantially. More importantly, domain-specific expertise can provide more accurate risk assessments of a megafund's holdings, reducing the element of surprise for investors, portfolio managers, and researchers.

The fact that such collaboration does not yet exist may be a symptom of a deeper divide between academia and the biopharma industry: a cultural gap between scientific research and commercial enterprise. In a comprehensive study of the business of science—with particular emphasis on biopharma—Pisano (2010) provides an eloquent summary of this gap:[24]

> Science is a world focused on "first principles" and methods; in contrast, business worries about commercially feasible products and processes. Science is inhabited by academics; the manager, the industrial scientists, and the engineer dominate business. Both science and business are intensely competitive worlds but their markets and currency are distinct. In science, score is kept by peer review and grant givers, and measured ultimately by reputation; in business, score is kept by capital markets and measured by profitability. Publication is synonymous with science, secrecy synonymous with business.

While the research-backed obligation structure with long-term debt relieves some of the exigencies of shorter-term funding, it does not address the fundamental conflict between science and business.

However, the new structure of a megafund presents an opportunity to re-engineer the business of biomedical science. With sufficient financial clout to invest for the long run and withstand economic downturns, but enough flexibility to change the composition of its portfolio in response

to new scientific breakthroughs or shifts in economic or political climate, the megafund may be the ideal balance between stability and agility. To achieve this balance, the corporate governance structure and investment process of the megafund must be carefully crafted to promote collective intelligence while maintaining focus on the overall purpose of easing the burden of disease. In particular, the objectives of a megafund would not be the same as those of grant-making agencies such as the National Institutes of Health whose primary focus is supporting basic scientific research (therefore, megafunds would be complementary, not competitive, to current government funding). Accordingly, its investment team must be staffed by a combination of industry professionals with scientific, engineering, and business expertise, and with access to a wide network of scientific advisors to serve both as consultants and talent scouts. Too much centralization and control can sometimes stifle creativity and independence, so these staffing decisions must be made carefully to yield sufficient diversity while operating as a coherent group. New business arrangements may also need to be crafted to support truly innovative research. For example, investing in early-stage research may take the form of royalty-sharing agreements with university technology-licensing offices in which younger academics working in key areas of research—even if they have no immediate interest in preclinical applications—are offered unrestricted research funding in exchange for a small percentage of any future royalties that may be derived from their work.

It is well known that the complexity of managing organizations grows more than proportionally with size,[25,26] which may explain the recent empirical evidence that smaller more-focused biopharma companies seem more efficient than big pharma, producing a comparable number of new molecular entities at the same rate but at lower cost.[27] An even more relevant "proof-of-concept" of the efficiency of a megafund vehicle is offered by Royalty Pharma, which currently manages $8 billion with a full-time staff of 19 individuals (though they employ a much larger network of biomedical experts as consultants). A megafund can be managed effectively with a much smaller staff than a large pharmaceutical company because of the nature of its business—investing in biomedical projects, not operating drug-development, marketing, and distribution facilities. At the same time, megafunds can benefit from the staying power associated with deep pockets. This dual nature of megafunds is one of its most significant features—its impact on the industry is greatly multiplied by financial leverage, not by number of employees or the size of its plant and equipment.

Financial size offers several distinct benefits. By raising a large pool of capital dedicated to eradicating disease, the biomedical megafund can significantly increase public awareness for both the burden of disease and the potential for its cures, allowing the fund to gather proportionally greater resources to achieve its mandate. These resources involve more than just capital, long horizons, and financial diversification. They also include: research synergies, efficiency gains, and greater collective intelligence among multiple R&D teams (who would otherwise be prevented from exchanging ideas if they worked for unrelated competing companies); centralized management of clinical trials and shared information about their outcomes (especially negative results, which are currently not reported anywhere[28]); complementary educational synergies (e.g., facilitating a larger pipeline of M.D./Ph.D.s); stronger political support due to higher visibility among voters (e.g., government guarantees, tax incentives); and greater drawing power for hiring leading experts.

This last feature of a biomedical megafund may be the most effective way to bridge the cul-

tural gap between scientists and business executives. The most talented biomedical researchers may not be motivated by financial gain. However, an opportunity to join an elite team of like-minded researchers, engineers, and clinicians devoted to a worthy humanitarian challenge—with a vast pool of capital at its disposal that is more patient than the longest-horizon venture-capital fund, and an organization focused on reducing the burden of a disease they care deeply about—may be considerably more compelling.[29] And with current challenges such as cancer, heart disease, dementia, Alzheimer's disease, diabetes, obesity, malaria, and influenza, there is no shortage of projects with great social significance to support several biomedical megafunds. Large-scale diversified drug-development efforts facilitated by megafunds not only increase the likelihood of success, but also increase the economic value of these enterprises to all stakeholders. With sufficient scale, it becomes possible to do well by doing good.

The combination of social relevance and the profit motive may seem confusing and inappropriate to some, but this trend is becoming more prevalent as we face societal challenges that require an unprecedented scale of collaboration among millions of individuals. Although charitable giving is an important part of translational medical research, the magnitude of such giving is dwarfed by the pool of investment capital seeking a reasonable rate of return. By creating financial incentives for solving social problems like cancer, society is able to tap into this much larger pool of assets. The megafund can be viewed as another example of the broader trend toward "venture philanthropy" as practiced by existing organizations such as the Gates Foundation (gatesfoundation.org), the Robin Hood Foundation (robinhood.org), and the Children's Investment Fund Foundation (ciff.org). Another form of this trend is public-private investment programs, in which private-sector institutions provide financing under certain types of government sponsorship. Such programs played an important role in dealing with the recent financial crisis by raising over $29 billion of investment capital to purchase distressed securities.[30] Several important government initiatives are already underway for speeding up translational medical research such as the U.S. government's National Center for Advancing Translational Sciences (which is part of the Cures Acceleration Network) and the Israeli Life Sciences Fund. But with budgets of only $575 and $200 million, respectively, these efforts will eventually also require substantial private-sector funding—megafunds may be one solution.

Government can play other important roles in addition to providing seed funding. The regulatory approval process for therapeutics is one of the most critical drivers of risk and return in the biopharma industry, and innovations in "regulatory science" can increase productivity and reduce risk. Such innovations include more effective use of clinical data and providing more detailed feedback to the industry regarding the approval process, and the U.S. Food and Drug Administration has undertaken several initiatives along these lines (see, for example, the academic Partnership in Applied Comparative Effectiveness Science).[31] The government can offer broader incentives for translational medical research by extending the patent life of therapeutics in certain high-priority and areas such as cancer and heart disease.[32] More targeted support can be provided in the form of tax incentives (e.g., allowing corporations to repatriate offshore assets at reduced tax rates if invested in designated public-private partnerships designed to support biomedical innovation), guarantees for investors' assets in biomedical megafunds, and the establishment of a blue-ribbon advisory panel of leading academics and business leaders to accelerate public-private investment partnerships in this industry.

Government involvement is also likely to be necessary because of ethical and humanitarian considerations, which affect the biopharma industry more directly than others. Financial innovations that explicitly address such concerns in advance may also be worth exploring.[33] For example, if a cure for a lethal type of cancer is developed, should its price be whatever the market will bear? What about rare/orphan and developing-country diseases? Government intervention would almost surely impose price limits in the former and subsidies in the latter because of the ethical considerations surrounding life-or-death choices linked to economic profit. The inevitable consequences of the political economy of public health suggest that pure profit-maximizing behavior in the life sciences industry may not be sustainable. One approach to addressing this issue is to incorporate broader social objectives directly into the capital structure of the megafund. For example, beyond a certain threshold of profitability, a portion of the megafund's excess profits might be used to subsidize drugs for those least able to afford them. The immediate effect of such corporate policies would likely be reduced upside potential for equity-tranche investors; no doubt, the market price of the research-backed obligations' equity tranche would adjust swiftly to reflect these new policies. However, if such policies lead to a larger and more sustainable enterprise and broaden the appeal of the securities issued under such terms to a wider investor population, the ultimate impact on the amount of capital raised and the megafund's likelihood of success may, in fact, be positive. This possibility bears further investigation.

Finally, in much the same way that securitization may have been too successful in raising large pools of capital for U.S. residential real estate, megafunds may also enjoy rapid success that brings a new set of challenges which should be anticipated and addressed in advance. Rules regarding sales practices, disclosure requirements, permissible corporate governance structures, and suitability criteria for investors must be imposed and strictly enforced to ensure that megafunds serve their purpose without jeopardizing the stability of the financial system. Although it is impossible to guarantee that megafunds will generate attractive returns, much can be done to ensure that all stakeholders are fully aware of its risks.

Supplementary Empirical Results

Data from the ThomsonOne database, VentureXpert (VX), indicates that over the last decade the biotech and healthcare venture capital (VC) investments have exhibited significantly lower returns than in the past. This pattern suggests that venture capital investment in this sector may be suffering from a secular downturn in returns. In VX, the biotech sector includes human therapeutic biotechnology, industrial biotechnology, and biosensors, and the medical/healthcare sector covers pharmaceutical research, therapeutics, diagnostics, and other healthcare related services. The VX database provides returns from all stages of venture investment, and reports 1-year rolling-horizon internal rates of return based on cash inflows and outflows in each year. Those returns include results from both active and liquidated funds (avoiding survivorship bias in the data), and are net of management fees and carried interests.

Figure 2 contains the 1-year IRR (the blue and green lines) and trailing 10-year IRRs (the orange lines) for the biotechnology and medical/healthcare sectors, where the 10-year IRRs are computed by compounding the 1-year IRRs over the preceding decade.

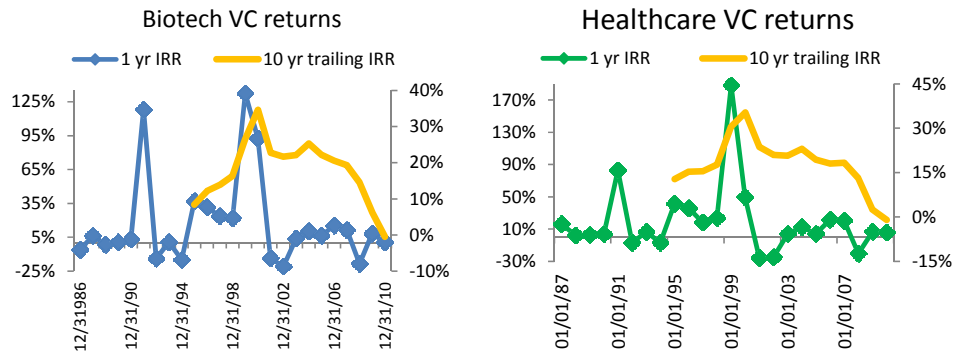

Figure 2: 1-year and trailing 10-year internal rates of return for the U.S. biotechnology and medical/healthcare sectors.

A recent paper by Booth et al. (2011)[34] also discusses the return of life sciences venture investing over the 2000–2010 period using information from a benchmarking database from Cambridge Associates (an investment advisory firm). Focusing on returns from deals (not funds), they examine the gross (including fees) pooled mean IRR for healthcare VC investing and its subsectors, and report a return of 15% for realized deals in the healthcare sector and 7.4% if unrealized deals are included.

Table 12 compares the 2000–2010 results from VX and Booth et al. (2011) (realized and unrealized deals) with the simulations results from our model. Using parameters derived from historical data on valuations from 2000 to the first quarter of 2011, the megafund yielded an annual return of 7.2% for both Simulations A and B with an all-equity capital structure. Simulations in which the capital structure consists only of equity are closest in structure to a large venture capital fund, hence a comparison of these results to those from VX and Booth et al. (2011) seems appropriate. The VX database shows an IRR of 2.7% for healthcare and 5.7% for biotech during 2000–2010. Since these returns are net of fees, we must add back the fees before comparing them to our simulation returns. The standard VC fees are approximately 2% per annum of assets under management plus carried interest (which we estimate to be an additional 1% per annum). This implies annual gross returns of 5.7% and 8.7%, respectively. The Booth et al. (2011) estimates of 12.8% and 7.6% already include fees. The megafund simulation yields net-of-fee returns of 7.2%, or 7.7% if we add the 0.5% service fees included in our simulation. Even though the assets involved in each of these cases do not match exactly, they are similar and, based on these comparisons, our simulation results seem consistent with recent historical experience in the biopharma industry.

Finally, it is important to note the downward trend in investment returns in the biotechnology and pharmaceutical venture capital sector over the past decade. In fact, if the year 2000—the last year the biopharma industry experienced large positive performance—is dropped from the sample, the IRRs for biotech and healthcare become negative according to VX (-0.5% for biotech and -0.7% for the medical/healthcare). This sensitivity to outliers suggests the importance of monitoring the investment performance of both sectors so as to recalibrate the simulations as needed.

Source	Raw	Gross of Fees
Simulated Megafund	7.2%	7.7%
VX (Healthcare)	2.70%	5.70%
VX (Biotech)	5.70%	8.70%
Booth et al. (2011) (All healthcare)	—	7.40%
Booth et al. (2011) (Pharmaceuticals)	—	12.80%
Booth et al. (2011) (Biotech)	—	7.60%

Table 12: Comparison of historical returns of VC equity investment from 2000–2010 and simulated megafund returns.

1. Higgins, R., Hamermesh, R. & Fuller, V. Note on the U.S. food and drug administration. *Harvard Business School* **9-807-050** (2007).

2. http://www.tamas.gov.il/NR/exeres/B7EF4EA0 2D41-4A97-8076-B080548A4990.htm.

3. Paul, S. M. *et al.* How to improve R&D productivity: The pharmaceutical industry's grand challenge. *Nature Reviews Drug Discovery* **9**, 203–214 (2010).

4. Bohn, J. & Stein, R. *Active credit portfolio management in practice* (Wiley Finance, 2009).

5. Dimasi, J. & Grabowski, H. Economics of new oncology drug development. *Journal of Clinical Oncology* **25**, 209–216 (2007).

6. Natanson, L. Innovative strategies for oncology drug development. Tech. Rep., Deloitte Recap LLC (2010).

7. Reichert, J. & Wenger, J. Development trends for new cancer therapeutics and vaccines. *Drug Discovery Today* **13**, 30–37 (2008).

8. Walker, I. & Newell, H. Do molecularly targeted agents in oncology have reduced attrition rates? *Nature Reviews Drug Discovery* **8**, 15–16 (2009).

9. Vasicek, O. Probability of loss on loan portfolio. Tech. Rep., KMV Corporation (1987).

10. Metrick, A. & Yasuda, A. *Venture Capital and the Finance of Innovation* (Wiley, New York, 2010), 2nd edition edn.

11. McCully, M. G. Finance and business development for biotechnology. Tech. Rep., UCLA Guest Lecture (Frontiers in Biotechnology) (2009).

12. Dimasi, J. A., Hansen, R. W. & Grabowski, H. G. The price of innovation: new estimates of drug development costs. *Journal of Health Economics* **22**, 151–185 (2003).

13. Light, D. W. & Warburton, R. Demythologizing the high costs of pharmaceutical research. *BioSocieties* **6**, 34–50 (2011).

14. Adams, C. P. & Brantner, V. V. Estimating the cost of new drug development: Is it really $802 million? *Health Affairs* **25, no.2**, 420–428 (2006).

15. Stewart, J. & Naeymi-Rad, N. Oncology national commercial study: Ten transformative trends (2011). URL `http://www.campbellalliance.com/landmark/index.cfm?sl_id=2`.

16. Huggett, B. Biotech's wellspring: A survey of the health of the private sector. *Nature Biotechnology* **30**, 395–400 (2012).

17. Cyranoski, D., Gilbert, N., Ledford, H., Nayar, A. & Yahia, M. The ph.d. factory. *Nature* **479**, 276–279 (2011).

18. Nanda, R. & Rhodes-Kropf, M. Financing risk and innovation (2011). URL `http://ssrn.com/abstract=1657937`.

19. Nanda, R. & Rhodes-Kropf, M. Investment cycles and startup innovation (2011). URL `http://ssrn.com/abstract=1950581`.

20. Hsu, P.-H., Tian, X. & Xu, Y. Financial development and innovation: Cross-country evidence (2012). URL `http://ssrn.com/abstract=1745682`.

21. MacConaill, L., van Hummelen, P., Meyerson, M. & Hahn, W. Clinical implementation of comprehensive strategies to characterize cancer genomes: Opportunities and challenges. *Cancer Discovery* **1**, 297–311 (2011).

22. Finkelstein, S. & Temin, P. *Reaonable Rx: Solving the Drug Price Crisis* (Pearson Education, Upper Saddle River, NJ, 2008).

23. Sharp, P. Q&A: Phillip Sharp on biomedical convergence. *Cancer Discovery* **1**, 370 (2011).

24. Pisano, G. P. The evolution of science-based business: Innovating how we innovate (2010). URL `http://www.hbs.edu/research/pdf/10-062.pdf`.

25. Perrow, C. *Complex Organizations: A Critical Essay* (McGraw-Hill, New York, 1986).

26. Anderson, P. Complexity theory and organization science. *Organization Science* **10**, 216–232 (1999).

27. Munos, B. Lessons from 60 years of pharmaceutical innovation. *Nature Reviews: Drug Discovery* **6**, 959–968 (2009).

28. Wood, A. J. Progress and deficiencies in the registration of clinical trials. *New England Journal of Medicine* **360**, 824–830 (2009).

29. Baylin, S. & Jones, P. Q&A: Stephen Baylin and Peter Jones on team science. *Cancer Discovery* **1**, 279 (2011).

30. U.S. Treasury. *Legacy Securities Public-Private Investment Program: Program Update – Quarter Ended March 31, 2012* (U.S. Treasury, Washington, D.C., 2012).

31. Food and Drug Administration. *Advancing Regulatory Science at FDA* (U.S. Food and Drug Administration, Washington, D.C., 2011).

32. Frost, D. A. Extend drug patent life to rejuvenate r&d. *Financial Times* **May 8** (2012). URL `http://www.ft.com/cms/s/0/4d670934-954a-11e1-8faf-00144feab49a.html#axzz24XI9AV4n`.

33. Frangioni, J. V. You can't give it away in the USA, or can you? Tech. Rep., Center for Molecular Imaging, Beth Israel Deaconness Medical Center (2012).

34. Booth, B. L. & Salehizadeh, B. In defense of life science venture investing. *Nature Biotechnology* **29**, 579–583 (2010).

Acknowledgements . We thank Jim Broderick for his advice and help in understanding the challenges of developing and financing new drugs, and for motivating our interest in this area in discussions with A.W.L. in 2007.

We also thank Janice Reichert and the Center for the Study of Drug Development (Tufts University School of Medicine) for generously sharing their data with us and for her guidance and support in accessing this data, and Lisa Natanson of Deloitte Recap LLC for giving us access to Recap's DEAL builder™ and DEVELOPMENT optimizer™ online tools which contained detailed information on the economics of licensing deals and clinical trials of oncology compounds.

Lloyd Han, James Noraky, Ashutosh Singhal, and Christopher Wilfong provided research assistance throughout the entire project, and we also acknowledge Allister Bernard, Helen Hway Chen, Ming Jack Chen, Rumela Das, and Roman Garcia for research support during various phases of the project. Finally we thank the editor, Kathy Aschheim, and three reviewers for constructive comments and editorial advice, and David Agus, Jim Broderick, Lewis Cantley, Shirish Chinchalkar, John Cox, Jayna Cummings, Ora Dar, Ashish Das, John Evans, John Frangioni, Al Gilman, Jacob Goldfield, Duncan Higgons, Tom Kalil, Eugene Kandel, Michael Kanef, Andrew Kimball, Doug Jamison, Eric Lander, Pablo Legorreta, Leonard Lerer, Judy Lewent, Francois Maisonrouge, Parag Mallick, Monique Mansoura, Nick Marko, Albert Metz, Bob Merton, Fiona Murray, Krishna Murthi, Larry Nagahara, Bob O'Neill, Averill Powers, Rory Riggs, Duane Roth, Maya Said, Melissa Stevens, Kailash Swarna, Tom Tombrello, Ivan Wolff, Alastair Wood, Glenn Yago, and participants at the Financial Innovations Lab Workshop hosted by the Milken Institute and FasterCures in July 2011, the 2012 Milken Institute Conference, and the 2012 USC NCI Emerging Approaches in Oncology conference for helpful comments and discussion.

The views and opinions expressed in this article are those of the authors only and do not represent the views and opinions of MIT, AlphaSimplex, Moody's Corporation, any of their affiliates or employees, or any of the individuals acknowledged above. Research support from the MIT Laboratory for Financial Engineering is gratefully acknowledged.

Author Contributions All authors contributed equally to this research. A.W.L. first developed the idea for securitizing biomedical research after conversations with Jim Broderick in March 2007 about a portfolio

approach to biomedical innovation. A.W.L. assembled key members of the project team, provided funding through the MIT Laboratory for Financial Engineering, and was responsible for overall project management. JM.F. was responsible for coordinating all aspects of the project, including directing research assistants, obtaining and processing all input data, calibrating the simulation parameters, running the simulations, and preparing the initial draft of the manuscript, with input and oversight from A.W.L. and R.M.S. R.M.S. developed the analytic framework for modeling the portfolio of drug compounds. R.M.S. and Lloyd Han developed the R code with assistance from James Noraky and JM.F, and input from A.W.L. and Ashutosh Singhal. Allister Bernard converted the R code to Matlab. A.W.L. and JM.F. validated the final version of the Matlab code. R.M.S. also prepared the description of the simulation results, which was reviewed and revised by JM.F. and A.W.L. A.W.L. constructed the illustrative portfolio example and prepared the final draft of the manuscript, with input and revisions from JM.F. and R.M.S.

Competing Interests JM.F. declares no competing interests.

R.M.S. declares the following competing interests: In addition to his MIT LFE position, R.M.S. is managing director at Moody's Corporation; member, Board of Directors, PlaNet Finance US; member, Advisory Council, Museum of Mathematics. President, Consortium for Systemic Risk Analytics.

A.W.L. declares the following competing interests: In addition to his MIT faculty position, A.W.L. is a Research Associate, National Bureau of Economic Research; Chief Investment Strategist, AlphaSimplex Group; consultant, Office of Financial Research; member, Moody's Advisory and Academic Research Committee; member, Financial Advisory Roundtable, Federal Reserve Bank of New York; member, Economic Advisory Committee, FINRA; member, Board of Overseers, Beth Israel Deaconess Medical Center; member, Academic Advisory Board, Consortium for Systemic Risk Analytics.

No funding bodies had any role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. No direct funding was received for this study; general research support was provided by the MIT Laboratory for Financial Engineering and its sponsors. The authors were personally salaried by their institutions during the period of writing (though no specific salary was set aside or given for the writing of this paper).

Correspondence Correspondence regarding the manuscript should be addressed to Andrew W. Lo, MIT Sloan School of Management, 100 Main Street, E62–618, Cambridge, MA 02142–1347, USA (email: alo@mit.edu). Correspondence regarding the simulation software should be addressed to Jose-Maria Fernandez, MIT Laboratory for Financial Engineering, 100 Main Street, E62–611, Cambridge, MA 02142-1347, USA (email: jose-maria.fernandez@sloan.mit.edu).