How to Remove Chronic Optimism from your Compound and Portfolio Evaluations while Creating More Value

When selecting compounds to advance in clinical trials, buying or selling an asset, sequencing trials or allocating resources, you rely on estimates of compounds' values. Why are high evaluations often optimistic? Why are portfolios perennially overvalued? Why do you suffer reoccurring disappointment as compounds and portfolios achieve fractions of their estimated values? Many managers blame the evaluation process, accusing it of optimism, but persistent optimism arises even when evaluation processes are unbiased. This paper explains why and predicts that, on average, unless you adjust them, forecasts of blockbuster sales are 100% too high and assessments of phase III portfolios are 40%-50% overvalued. I present a solution to this problem, one developed by management scientists, and used by executives who manage oil and gas exploration. Importantly, the fix boosts productivity as well, possibly raising portfolio value by 10% or more. This paper eschews mathematics to focus on explanation and understanding.

Gary J. Summers, PhD
Pipeline Physics LLC
www.PipelinePhysics.com

How to Remove Chronic Optimism from your Portfolio Evaluations while Creating More Value

When allocating resources, sequencing clinical trials, selecting compounds to advance downstream, buying and selling assets, managing portfolio risk or maximizing portfolio value, do you forecast the revenues or profits compounds might produce if they survive development? How do errors in these forecasts affect your decision-making and results? How can you improve your decisions and create more value?

I am not a pharmaceutical executive, and the ideas I present below come, not from experiencing your responsibilities, but from a portion of probability theory called the algebra of random variables. The theory's prognosis is cogent and consequential.

1. On average, forecasts of high net present value (NPV) are optimistic, and this optimism increases with the size of the forecast. A drug in development that is predicted to be a blockbuster (if it launches), on average, produces only half of its estimated profits.

2. NPV and expected NPV (eNPV) consistently overestimate portfolio value. For a portfolio of phase III compounds, the overestimate is commonly 40%-50%.

3. The errors in forecasts of NPV and eNPV harm project selection and reduce productivity by at least 15%. This reduction occurs whether you use simple selection methods, like rankings and cutoffs, or sophisticated ones, like portfolio optimization.

Fortunately, a solution exists:

4. Adjusting NPV estimates by using information about classes of compounds, derived from historical data, eliminates the first two problems and reduces the third one by 50%.

These implications arise from three qualities of drug development: (1) the NPVs of drugs are highly skewed, (2) NPV forecasts are imprecise and (3) forecasting errors scale with profits. You are not alone in suffering the consequences of these characteristics. The executives who guide oil and gas exploration face them as well, and they have a solution you can use (Begg and Bratvold 2008; Schuyler and Nieman 2008; Chen and Dyer 2009).

I introduce each of these qualities, show how they create the above problems and then present the solution to you. The presentation builds intuition and understanding, and if you desire, you can find the mathematics in my academic paper (Summers 2018).

The skewed distribution of NPVs: Pharmaceutical profits are highly skewed. Grabowski et *al*. (2002) report that 10% of drugs produce 50% of the industry's profits, while 30% of compounds

yield 80% of the profits. Figure 1 presents a probability distribution matching these qualities. Its horizontal axis is in units of the average NPV, denoted as $\overline{NPV}$. For example, if you are managing the development of oncology compounds, $\overline{NPV}$ represents the average NPV of an oncology drug.

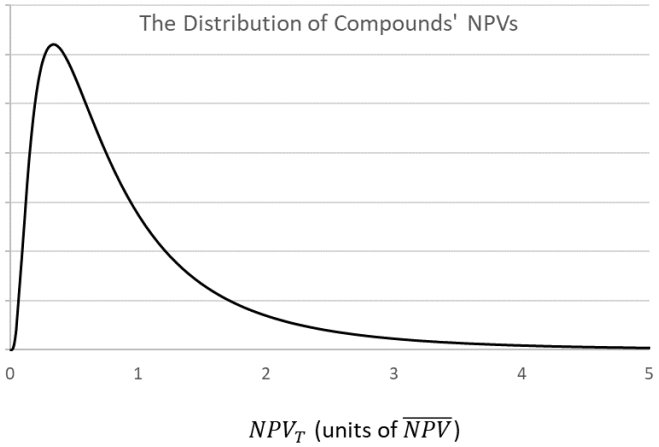


Figure 1: A highly skewed distribution that illustrates the profits produced by pharmaceuticals, reported by Grabowski et al. (2002). The variable NPV_T stands for true NPV, as opposed to forecasts, which I denote with NPV_F. The horizontal axis is in units of average NPV, $\overline{NPV}$, so the average (expected value) of the distribution occurs at unit one. In this dispersal, 74% of compounds have below average NPV, and 12% of compounds have values greater than twice the industry average.

Highly imprecise forecasts: By analyzing historical data, Cha et *al*. (2013) estimate the error in predictions of peak sales. On average, forecasts made two years before launch have errors equaling 75% of the peak sales the drugs achieved. Made when phase III was complete, or nearly so, these forecasts exploited strong knowledge of safety and efficacy. In contrast, when you select phase II compounds to advance, you face more uncertainty, especially about patient populations, endpoints, biomarkers, dosages, safety, and efficacy. For these crucial selection choices, NPV forecasts may have errors equaling or exceeding 100% of the NPV a drug will produce if launched. Figure 2 illustrates a distribution of forecasting errors with an average error (standard deviation) of 100%. Notice how the distribution centers on a compound's true NPV, NPV_T, so the errors are unbiased. They overestimate NPV_T with the same frequency and size as they underestimate it. You could add bias to this model, but I wish to show how unbiased forecasting errors produce optimistic estimates of NPV and portfolio value (problems 1 and 2 above).

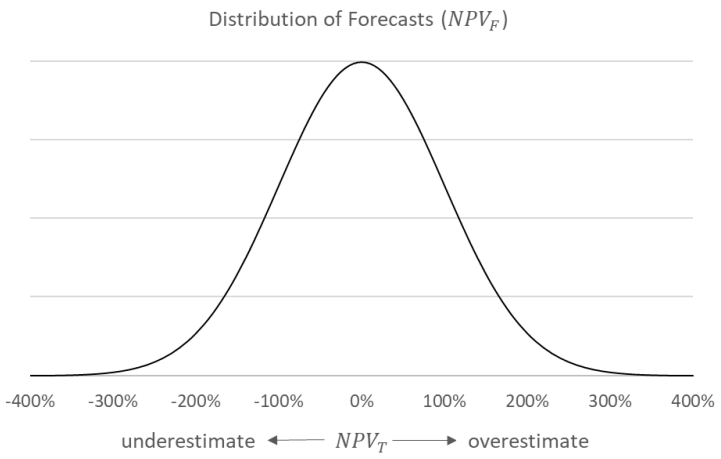


Figure 2: This distribution of compounds' NPV forecasts (NPV_F) is unbiased and centered on the true NPVs (NPV_T). The horizontal axis shows the errors in the forecasts, with deviations measured as a percentage of NPV_T. Positive deviations are overestimates; negative deviations are underestimates. The standard deviation of the errors is $100\%NPV_T$. The errors are normally distributed, so about 68% of the forecasts are within $\pm 100\%NPV_T$ of their true values and 95% of the forecasts are within $\pm 200\%NPV_T$ of their true values.

Errors scale with NPV: Suppose a drug has a small market and a second drug has a large one. Measured in absolute size, forecasting errors are lesser for the small market. Mathematically, you can model the forecasted NPV (NPV_F), as arising from the true NPV (NPV_T) and a forecasting error: $NPV_F = NPV_T(1 + \varepsilon)$, or equivalently, $NPV_F = NPV_T + \varepsilon NPV_T$. The variable ε determines the forecasting error. If $\varepsilon = 50\%$, the forecasting error is $50\%NPV_T$. Recall that two years prior to launch, the average error, which is also called the standard deviation, is 75%. This error, which is also called imprecision, is reported in two ways: $SD(\varepsilon) = 75\%$ and Error $= \pm 75\%NPV_T$.

How can unbiased errors consistently overestimate profits and portfolio value? The phenomenon is called the optimizer's curse (Smith and Winkler 2006), and it afflicts all project selection methods, including simple techniques, like rankings and cutoff values, and sophisticated ones, like optimization.

Problem 1: Above average forecasts of NPV are (on average) optimistic

Are you interested in compounds with above average value? Often, predictions of superior worth are optimistic, and these rose-colored valuations harm your portfolio management,

including your risk management, selection and canceling of compounds, sequencing of clinical trials and allocation of resources.

To see how optimism arises, consider these statements:

A) Most graduate students eat Ramen noodles.
B) Most people who eat Ramen noodles are graduate students.

Statement (B) switches the ordering of the nouns, from graduate students preceding Ramen noodles to Ramen noodles preceding graduate students. Obviously, statement (B) is incorrect. Reversing the order creates an error.

Now consider these statements:

C) An NPV forecast is equally likely to be above or below the true NPV.
D) The true NPV is as likely to be above or below the forecast.

Traversing from statement (C) to (D) creates the same logical error as in the Ramen noodles example, but to understand why we need some details. Think about Figure 2 and its unbiased errors. It implies statement (C), but only in two circumstances. First: you randomly select compounds to buy or advance downstream. Of course, instead of rolling dice, you strive to choose the best compounds, a subject we address in the next section. Second: you have a drug in development, are going to forecast its NPV, but you have not yet created the forecast. Consider this case.

Before making a forecast, if the assessment process is unbiased, overestimates are as likely as underestimates, but once created, the result is only one of these possibilities. Fortunately, you have information to indicate which one: the forecast itself. Suppose a compound in drug development is predicted to produce blockbuster sales (if it survives development). Likely, the stellar estimate arose in one of two ways: (1) the compound is indeed a blockbuster and the forecasting error is small or (2) the compound is not a blockbuster and the forecasting error is optimistic. Look at Figure 1. Few compounds are blockbusters, so the second situation is more likely. Despite accurate (unbiased) evaluations, forecasts of blockbuster NPVs are, on average, optimistic.

An analogy illustrates this logic. Suppose your scientists develop a diagnostic test for a disease that afflicts one of every one hundred people. The test has a negligible false-negative rate and a false-positive rate of 5%. In a hundred tests, you expect to get one true-positive and five-false-positives, so a person diagnosed with the disease has only a 17% chance (1/6) of having the

disease. Because the disease is rare, a positive test is likely to be a false-positive. Forecasts of blockbuster NPVs are analogous.

How big is this problem? Figure 3 presents the results of combining Figures 1 and 2 with Bayes' law (which is the correct way to switch the order of the nouns in statements A through D). Look at a forecast of five times the average profit ($5\overline{NPV}$), which is a good definition of a blockbuster (Grabowski et *al*. 2002). Figure 3 estimates that, on average, compounds with such glorious assessments produce only 45% of their predicted profits (problem 1 of this paper's introduction).

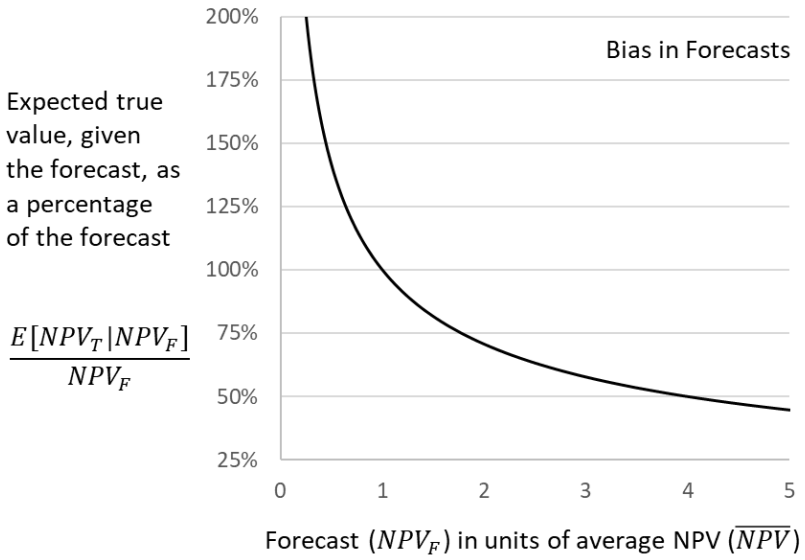


Figure 3: The graph shows the bias in forecasts. The horizontal axis is the forecasted NPV (NPV_F), measured in units of the average NPV of compounds ($\overline{NPV}$). The vertical axis is the expected true NPV (NPV_T), given the forecast, divided by the forecasted NPV. This ratio presents the expected true value as a percentage of the forecast. Forecasts of average value, $NPV_F = \overline{NPV}$, are unbiased. On average, predictions of above average value are optimistically biased, while estimates of below average value are pessimistically biased.

Figure 3 reveals the following: On average, forecasts of above-average NPV are optimistic, as are eNPVs and ROIs made with these forecasts. Treating such predictions as unbiased will cause mistakes when managing clinical trials and drug development portfolios.

An additional consequence is subtle. Suppose an expert tells you that a specific forecast is unbiased because he or she used debiasing methods when making the prediction. This statement may be frustrating to a pharmaceutical executive, especially after experiencing numerous optimistic evaluations. The expert's claim is erroneous. Debiasing techniques used while evaluating assets yield unbiased errors, like Figure 2, and therefore infer statement (C). Claiming a specific evaluation is unbiased commits the logical mistake of assuming statement (C) implies (D). In contrast, successful debiasing techniques, including the one presented herein, modify an estimate after you make it.

Problem 2: Portfolios are consistently overvalued (the optimizer's curse)

Are you routinely disappointed by underachieving portfolios? Do you feel exasperated when a decision scientist blames poor results on chance, saying that, while portfolio optimization maximizes value, lousy luck can produce bad results? You are right to be frustrated. You are not suffering ill luck. You are experiencing the optimizer's curse, a systematic, optimistic bias caused by project selection.

The problem is pervasive, affecting all nonrandom selection methods. In management science, it is called the optimizer's curse. In economics, it is called the winner's curse. In finance's modern portfolio theory, it is the impact of estimation errors. In statistics and machine learning, it is called overfitting. Focusing on drug development, whether you select compounds by group discussion, project rankings, cutoff values, portfolio optimization or another method, you suffer the optimizer's curse and are thus routinely disappointed by portfolios that produce less value than predicted.

When introducing the curse to management scientists, Smith and Winkler (2006) offered compelling examples. Suppose you select one of two projects for investment. The assets have the same expected value, but this fact is unknown to you. To make your choice, you first estimate their NPVs. While accurate (unbiased), your estimates are imprecise. Each assessment has a 50% chance of being optimistic and a 50% chance of being pessimistic. There are four possible outcomes:

- You underestimate the value of both projects (25% chance)
- You overestimate the value of one project and underestimate the value of the other one (50% chance)
- You overestimate the value of both projects (25% chance)

You select the project with the highest estimated value, so 75% of the time you overestimate the worth of your investment.

Now suppose you select one of three projects, and to present a vivid case, they are worth the same: $100 million. Your evaluation of each project is accurate (unbiased), but imprecision is 75%, the same average error Cha et *al*. (2013) discovered when studying the evaluations of compounds finishing development. Figure 4 shows the results. The solid curve is a probability distribution of the possible evaluations of each project. The dashed curve is a probability distribution of the highest estimate, and its mean is $163 million. Even though the valuations are unbiased, your assessment of the compound you choose overstates value, on average, by $63 million.

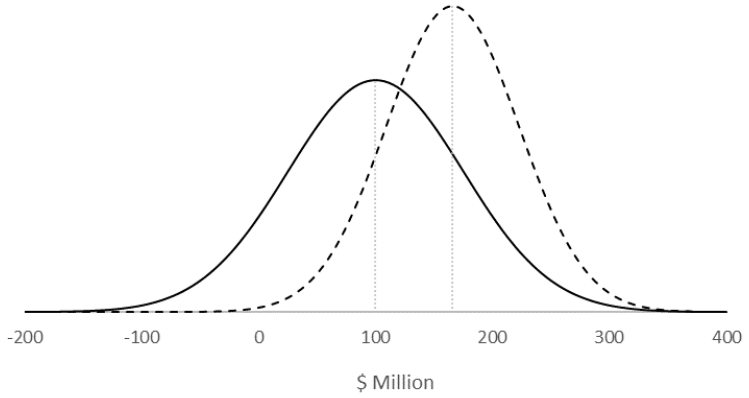


Figure 4: The solid curve is a probability distribution that shows possible eNPV valuations for each of three projects. The evaluations are imprecise (the standard deviation of the curve is $75 million), but the errors are unbiased. The dashed curve is a probability distribution of the maximum of the three estimates. Its expected value is $163 million. On average, you overestimate the worth of the selected project by 63%.

To see how the optimizer's curse affects drug development look at Figure 3. If you rank projects by their NPV forecasts and fund down the ranking, your choices will contain more overestimate than underestimate, unless you fund 100% of compounds. You will overvalue your portfolio and subsequently suffer disappointment. This disenchantment is the optimizer's curse.

Using Monte Carlo analysis, I estimated the overvaluation suffered when selecting phase II compounds to advance downstream (Summers 2018). I assumed the forecasting errors of Figure 2 and an exponential distribution of NPV, which while highly skewed, is less skewed than Figure 1. Using a simple model of a two-armed parallel group trial, I randomly assigned to each compound a probability of surviving phase III, denoted by PTS. Finally, I ranked projects by their forecasted expected values, $eNPV_F = PTS * NPV_F$. I assumed all compounds incur the same cost of phase III trials, so the ranking is equivalent to ordering assets by ROI or another

bang-for-the-buck metric. Figure 5 presents the results. It shows the overestimate in portfolio value as a function of phase II's success rate (throughput). Industry-wide, 34% of phase II compounds advance (Hay et *al*. 2014), and for this throughput, Figure 5 estimates that phase III portfolios are overvalued by 45% (problem 2).

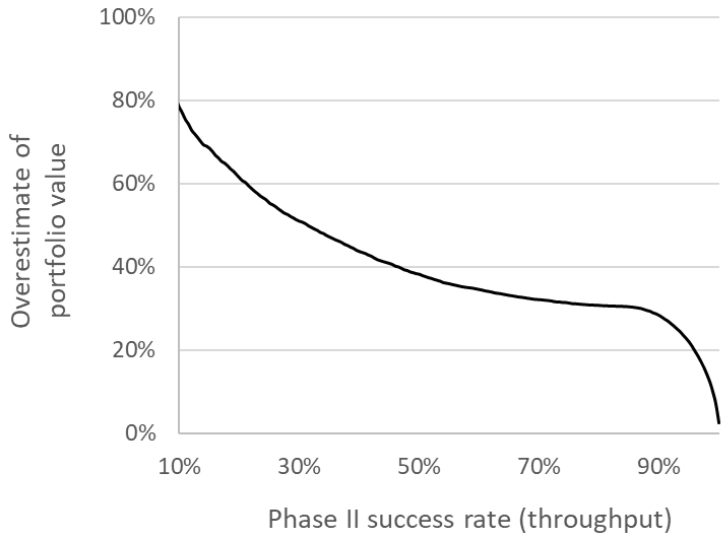


Figure 5: At the gate to phase III, overvalued compounds are more likely to advance than undervalued compounds. The curve shows the resulting overestimate in portfolio value as a function of phase II's success rate.

<u>Problem 3: Imprecise profit forecasts reduce portfolio value by at least 15%</u>

Overestimating compounds' and portfolios' values are secondary effects of a pernicious problem. Imprecise NPV estimates depress productivity.

To estimate the loss, we must define selection errors. A market false-positive is advancing an unprofitable compound, and a market false-negative is canceling a profitable one. A technical false-positive is advancing an unsafe or ineffective compound, and a technical false-negative is canceling a safe and effective one. We will focus on market selection errors.

For a specific compound, we would define profitability by considering its cost, but because our analysis occurs at the phase level, we will use averages. For compounds being considered for phase III trials, historical data (Grabowski et *al*. 2002; DiMasi et *al*. 2003) exposes the expected cost of creating one new drug to be $45\%\overline{NPV}$ (Summers 2018). A profitable compound produces profits that exceed this investment, $NPV_T > 45\%\overline{NPV}$, while an unprofitable one

yields less, $NPV_T \leq 45\%\overline{NPV}$. For the distribution of profits in Figure 1, 48.5% of compounds at the gate to phase III are profitable.

Now suppose you forecast NPV for each compound, set a cutoff value and advance only those compounds with estimates that exceed the cutoff. If your assessments were errorless, you would set the cutoff value at $45\%\overline{NPV}$, advance every profitable compound, cancel every unprofitable one and thereby maximize value. Unfortunately, forecasts contain large errors, so some unprofitable compounds have high evaluations, some profitable ones have low evaluations, and however you set the cutoff, you will commit some selection errors.

Figure 6: The percent of selection decisions that are errors when ranking compounds by forecasted NPV (NPV_F) and funding down the ranking. The curves arise from the distribution of NPVs in Figure 1 and the forecasting errors of Figure 2. At a minimum, 30% of decisions are selection errors.

Assuming Figure 1's distribution of value and Figure 2's estimation errors, which reasonably characterize drug development, Figure 6 illustrates the selection errors that occur when advancing phase II compounds. With 100% throughput, no compound fails. False-negatives are absent, but 51.5% of decisions are market false-positives. With no throughput, market false-positives are absent, but 48.5% of decisions are market false-negatives. In between, adjusting

the cutoff value trades false-positives for false-negatives. Because of imprecise forecasts, at least 30% of decisions are selection errors.[1]

Studies of clinical trials often present selection errors as rates. The market false-negative rate is the percentage of profitable compounds that executives cancel, while the market false-positive rate is the percentage of unprofitable compounds that executives advance. For phase II's 34% success rate, Figure 6 indicates a market false-negative rate of 47% and the market false-positive rate of 17%. Also, 75% of the advancing compounds are profitable, up from the 48.5% of phase II compounds.

You wish to maximize value, not to minimize the number of selection errors. Advancing a slightly unprofitable compound is hardly harmful, but canceling a blockbuster is disastrous. Suppose you estimate each compound's value, NPV_F, and its probability of technical success, PTS. Here are two methods of choosing compounds for phase III:

- $eNPV_F$: You multiply each compound's forecasted NPV by its probability of technical success to create an expected NPV, $eNPV_F = PTS * NPV_F$. Then you set a cutoff value and advance only those compounds with $eNPV_F$ above the cutoff. This is a compensatory selection method because a high value of one metric can compensate for a low value of the other metric. (Decision trees more accurately fold costs into values, but this simpler procedure is fine for my analysis.)

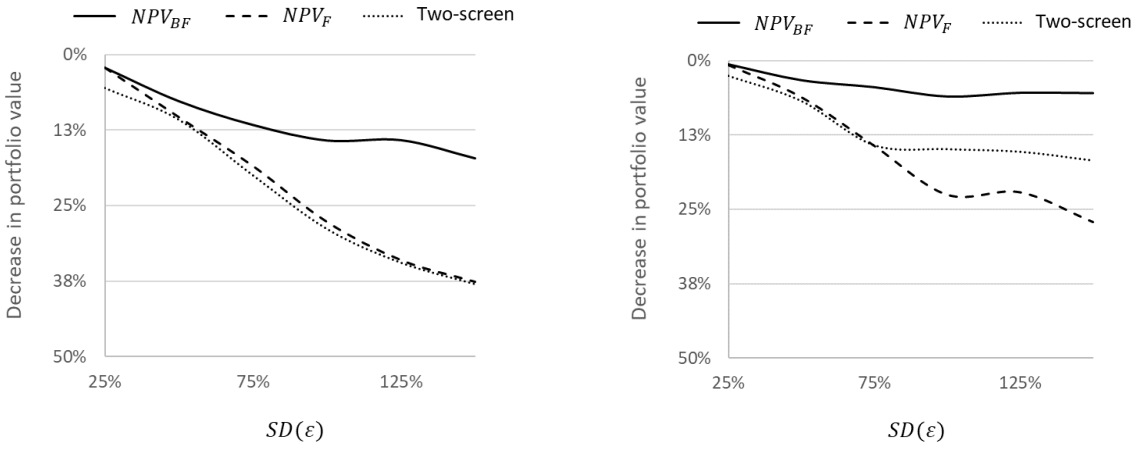
- Two-screen cutoff: You set two cutoff values, one for PTS and another for NPV_F, and advance only those compounds with estimates that surpass both cutoffs. This is a noncompensatory approach. A low value on either metric cancels a compound, regardless of the value on the other metric. This two-screen method is like the fast and frugal heuristics that beat "optimal" selection rules at some classification tasks (Katsikopoulos 2011).

Can the two-screen cutoff beat $eNPV_F$? Recall, reasonable estimates of the market false-negative and false-positive rates, based on historical data, are 47% and 17%. Meanwhile, common goals of phase II trial designs are technical false-negative and false-positive rates of 20% and 5% (Lindborg et *al*. 2014). The market error rates are much higher, suggesting that drug development faces more market uncertainty than technical uncertainty. Perhaps, multiplying the NPV forecasts by the probabilities of technical success corrupts the good technical information and thereby causes numerous selection errors. By keeping the information

[1] Because compounds with small true NPVs has small errors, the model $NPV_F = NPV_T(1 + \varepsilon)$, produces unrealistic behavior when throughput nears 100%. To avoid this problem, I made Figure 6 by using a lognormal model (Chen and Dyer 2009). For the same level of market uncertainty, the lognormal model sets the minimum error at 30%, while the original model sets the minimum at 29%.

separate, the two-screen method may commit fewer errors, and perhaps, create more value as well.

I tested this hypothesis with Monte Carlo analysis (Summers 2018). Selecting compounds based on their true expected values, $eNPV_T = PTS * NPV_T$, produces the maximum value available in a scenario. My study measured the amount of available value lost to imprecise forecasts when selecting compounds by $eNPV_F$ and the two-screen cutoff. To measure the minimum loss these techniques can produce, I used the optimal cutoff values for each method. Figure 7A shows a scenario where development costs, forecasting errors and the distributions of PTS and NPV_T match common values for the selection gate that precedes phase III. The $eNPV_F$ method slightly outperforms the two-screen cutoff, and while not shown in the figure, it is more robust as well, beating the two-screen approach for a wide range of cutoff values. Although $eNPV$ is superior, its advantage is small. If the two-screen method helps executives achieve consensus or justify their decisions, using it is reasonable.



Figures 7A (left) & 7B (right): Estimated via Monte Carlo analysis (Summers 2018), these graphs show how imprecise NPV forecasts affect the value of a phase III portfolio. The curves show the value lost by three selection methods: (a) expected NPVs made from Bayesian adjusted forecasts ($eNPV_{BF}$), (b) expected values made from traditional NPV forecasts ($eNPV_F$) and (c) the two-screen cutoff method. The horizontal axis is the size of forecasting errors, $SD(\varepsilon)$. Recall, at the gate to phase III, the errors are $SD(\varepsilon) > 75\%$. The charts represent two scenarios: the high cost of drug development, $50\%\overline{NPV}$ (7A), and a lower cost, $27\%\overline{NPV}$ (7B), that might characterize other industries.

Figure 7B presents the results for Figure 7A's scenario but with development costs decreased by almost half. As imprecision increases, the two-screen cutoff outperforms $eNPV_F$, thereby demonstrating the above hypothesis. In industries with highly skewed profits, development cost below those of pharmaceuticals, which characterizes most industries, and large forecasting errors, preferring the two-screen cutoff to $eNPV_F$ is rational.

Returning to the question that began this section: How much productivity is lost because of imprecise forecasts of NPV? Figure 7A shows that errors typical of drug development ($SE(\varepsilon) > 75\%$), reduce productivity by at least 15% (problem 3).

In one sense, the comparison is unfair. I prevented the two-screen cutoff from exploiting its most significant advantage – using any metric of market potential, not just NPV. Available information may reliably estimate the maximum profit a drug could produce. A high maximum is a necessary, but not sufficient, condition for substantial profits, so this metric can eliminate unprofitable compounds while creating few market false-negatives. It could be a useful screen, especially upstream in development and discovery.

Notice the solid lines in Figures 7A and 7B. They display the benefits of the technique presented below. This technique reduces the productivity loss by half, and it removes all biases in estimates of compounds' and portfolios' worth. You will see this solution next.

A solution to all three problems

We must defeat three defects: (1) biased forecasts, (2) overvalued portfolios and (3) productivity lost to imprecise estimates. Smith and Winkler (2006) proposed a strategy that fixes the first two flaws while mitigating the third.

Momentarily ignoring its details, applying their strategy to pharmaceuticals simple. You begin by creating your ordinary forecast, and then you define a reference class. Every compound is a member of a technical category, such as a therapeutic area (cardiovascular) or a mechanism (statins) and a market category, such as a novel drug, label expansion or me-too compound. Form these categories, identify the reference class that best exemplifies the drug, but define the class broadly enough to contain many members. Then derive Figures 1 and 2 (in mathematical form) from the class's historical data. Finally, combine your original forecast with Figures 1 and 2 via Bayes' law.

Abstracting further, Figure 1 presents data about a class of compounds, while your typical NPV forecast produces data about a member of this class, called a case. Having both case and class data is powerful. Studies by Nobel laureates Daniel Kahneman and Amos Tversky, and scientist Dan Lovallo, find that case estimates are habitually optimistic. Especially for projects, case

estimates overestimate value while underestimating cost and risk (Kahneman and Tversky 1982; Kahneman and Lovallo 1993; Lovallo and Kahneman 2003). These scholars recommend assessing a project's qualities by adjusting the case estimate with class data. Smith and Winkler (2006) implement this advice with Bayes' law. Their technique removes biases from forecasts, dispelling the optimizer's curse, to correctly value portfolios, while reducing forecasting errors, improving selection and raising productivity.

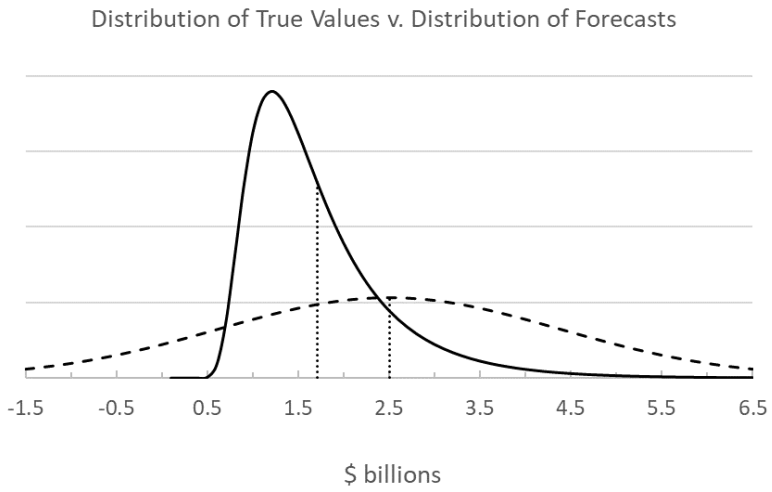


Figure 8: A forecast estimates a compound's value to be \$2.5 billion. Because the estimate is imprecise and above the average value ($\overline{NPV}$), it is likely to be optimistic. The solid line is a probability distribution of the compound's true NPV, given the forecast. It combines the forecast (case data) with Figure 1 (class data) while accounting for the forecast's imprecision ($SD(\varepsilon) = 75\%$). The distribution's mean, \$1.7 billion, estimates value without bias, and using this estimate minimizes evaluation errors, which improves selection and creates value. The dashed line shows a range of possible forecasts that a sensitivity analysis might produce. It is centered on the forecast and has an error of $\pm 75\% NPV_F$. Sensitivity analysis poorly approximates the solid line, and decisions based on its results will err.

Here is an example. Suppose you estimate a compound's NPV to be \$2.5 billion, with imprecision of $SD(\varepsilon) = 75\%$. The compound's reference class has an exponential distribution of value (somewhat less skewed than Figure 1) with an average NPV of \$1 billion. Integrating these qualities via Bayes' law produces the solid curve in Figure 8. It is a probability distribution that shows possible true NPV's, given your forecast. Notice its shape. The true NPV could exceed the forecast, but most likely, it is inferior. This distribution's average (expected) value is an unbiased estimate of the drug's true NPV. It is \$1.7 billion, 32% below the original forecast.

Using this estimate eliminates bias and minimizes forecasting error. If you apply the procedure to all your compounds, you will exercise optimism from your portfolio and value it correctly. By minimizing error, you will improve selection and create more valuable portfolios as well.

Sometimes, using expert opinion or sensitivity analysis, analysts will propose a range of estimates of a compound's value. Figure 8's dashed line illustrates one possible range, centered on the forecast, with an error of $\pm 75\% NPV_F$. By comparing the dashed and solid lines, you can see the forecast's optimistic bias and the range's incorrect estimate of imprecision. The mismatch arises because the dashed line presents a distribution of forecasts, but the solid line presents a distribution of true values. The dashed line answers the question: What are the possible forecasts of this compound's NPV? The solid line answers the question: Given our forecast, what are the possible true values of this asset? Presenting the dashed line as an estimate of the true value makes the same mistake as claiming statement C implies statement D. An expert in decision analysis would distinguish the lines of Figure 8 as follows: sensitivity analysis creates the dashed line; the solid line requires a value of information calculation.

Sensitivity analysis can help you improve forecasting as follows. If historical data is unavailable for estimating Figure 2, you can estiamte it with sensitivity analysis. However, you must avoid underestimating the imprecision (Russo and Schoemaker 1992).

To measure the benefits of the Bayesian adjustment, I tested it in the aforemented Monte Carlo analysis. For thousands of randomly created forecasts and errors, which produce NPV_F, I applied the Bayesian technique to create new estimates, NPV_{BF}. Then I created Bayesian adjusted expected values, $eNPV_{BF} = PTS * NPV_{BF}$ and selected compounds with the optimal cutoff value for this metric. Figure 7A reveals the results. For the imprecision of pharmaceutical forecasts, for drugs in development ($SD(\varepsilon) > 75\%$), the Bayesian adjustment reduced the productivity lost by about 50%. As Figure 7B shows, when development costs are smaller, the improvement is even greater.

Why is the Bayesian technique successful? Being compensatory, expected value trades probability of technical success for NPV and vice versa. Without the Bayesian adjustment, too much of the trade-off is erroneous, sacrificing some probability of success for forecasting error, for the last term of $NPV_F = NPV_T + \varepsilon NPV_T$. This erroneous trade-off explains why the performance of $eNPV_F$ degrades sharply as forecasting errors increase. The Bayesian modification reduces errors, which mitigates the problem.

As a general strategy for decision-making, combining class and case data integrates two distinct perspectives while infusing more information into your estimates. It resolves uncertainty.

Conclusion

Throughout my career, experts in decision analysis, whom I admire, advised me to distinguish good decisions from good results. They claimed that good decisions maximize expected value, but that bad luck can cause lousy outcomes. This reasoning is risky. It invites people to blame poor results on chance, instead of studying decision methods to find flaws and opportunities for improvement. Business situations are too multifaceted for any rational person to believe that all aspects of it are considered, thereby leaving luck as the sole source of poor results. Opportunities to improve decision-making always exist.

In drug development, an underperforming portfolio might be dismissed as a misfortune, perhaps because outcomes occur years after decisions, which makes learning hard. However, drug development portfolios consistently underperform, not because of chance, but because of the optimizer's curse – selection based on imprecise metrics creates optimism. Unless you adjust NPV forecasts, your portfolios will routinely disappoint you. The Bayesian adjustment solves the problem by mining historical data, and it thereby exemplifies the analytic innovations proliferating in many industries.

Recall the three previously mentioned selection criteria: two-screen cutoff, expected value, and Bayesian adjusted expected value. The two-screen cutoff provides a method of managing uncertainty, and in some situations, it beats expected values built from traditional forecasts. However, the two-screen cutoff gains scant benefit from the Bayesian technique, because for compounds in the same reference class, it leaves rankings unchanged. In contrast, expected values benefit mightily from the Bayesian adjustment, and these modified values produce the best selection. The implication is fundamental. Accommodating uncertainty is less effective than resolving it.

References

Begg SH, Bratvold RB (2008) "Systematic prediction errors in O&G project and portfolio selection," *Society of Petroleum Engineers, Annual Technical Conference and Exhibition*, Sept. 21-24, Denver, CO.

Cha M, Rifia B, Sarraf P (2013), Pharmaceutical forecasting: throwing darts? *Nature Reviews Drug Discovery*, 12(10): 737-738.

Chen M, Dyer J (2009), Inevitable disappointment in projects selected on the basis of forecasts, *Society of Petroleum Engineers Journal*, 14(2): 216-221.

Grabowski HG, Vernon J, DiMasi JA (2002) Returns on research and development for 1990s new drug introductions. *PharmacoEconomics* 20(Suppl. 3): 11-29

Kahneman D, Lovallo D (1993) Timid choices and bold forecasts: a cognitive perspective on risk taking. *Management Sci*. 39(1): 17-31.

Kahneman D, Tversky A (1982) Intuitive prediction: biases and corrective procedures. Kahneman D, Slovic P, Tversky A, eds. *Judgment Under Uncertainty: Heuristics and Biases* (Cambridge University Press: USA), 414-421

Katsikopoulos KV (2011) Psychological heuristics for making inferences: definition, performance and the emerging theory and practice. *Decision Analysis*. 8(1): 10-29.

Lindborg SR, Persinger CC, Sashegyi A, Mallinckrodt C, Ruberg SJ (2014) Statistical refocusing in the design of phase II trials offers promise of increased R&D productivity. *Nat. Rev. Drug Discov*. 13(8): 638-640.

Lovallo D, Kahneman D (2003) Delusions of success – how optimism undermines executives' decisions. *Harvard Bus. Rev*. 81(7): 56-63

Russo JE, Schoemaker PJH (1992) Managing overconfidence. *Sloan Management Review*, 33(2): 7-17.

Schuyler J, Nieman T (2008), Optimizer's curse: removing the effect of this bias in portfolio planning, *Society of Petroleum Engineers: Projects Facilities Construction*, 3(1): 1-9.

Smith JE, Winkler RL (2006), The optimizer's curse: skepticism and postdecision surprise in decision analysis, *Management Science*, 52(3): 311-322.

Summers GJ (2017), "Imprecise profit forecasts and the productivity of project selection criteria," Pipeline Physics LLC working paper.