TB Drug Development Portfolio Figures

Andrew Farlow
Department of Economics, and Oriel College,
University of Oxford
1 May 2006

The interest in taking an independent opinion on TB drug development portfolio figures was driven by the recent suggestion in a paper in the journal Science, Glickman et al.[1], that the current TB drug development pipeline and levels of funding are unlikely to yield a novel TB drug by 2010, and that achieving 95% confidence of at least one successful new drug would require as much as $400m in clinical trial costs and take 12 years. Hopefully this briefing note will help to clarify the situation.

1) Glickman et al. make use of Monte Carlo techniques to simulate various TB drugs portfolios. Intuitively, this process involves taking repeated drawings from various data distributions (trial transition probabilities, costs, trial duration, etc), and deriving distributions and relevant confidence intervals for the expected costs of development of TB drugs by a particular date.
    a. This is an appropriate method to use. The results do however depend on the original 'transition success probabilities' being appropriate for the problem at hand and independent, and on the trials being sequential, and on the accuracy of the underlying cost data and rates of discount used, etc. The results of such an exercise may appear more 'scientific' than is actually the case given that the results of such simulations are only as good as what gets fed into them. It is best to treat such simulations as a useful exercise with caveats.
    b. The Science background paper says that two hundred simulations were done to get within 1% of the "real" success probabilities. However, these are not the 'true' probabilities in the real world per se, but are instead based on the underlying probabilities derived from a consultation process. The Monte Carlo is random over this probability space, but this space itself is based on these estimates of transition probabilities, and therefore only as good (or 'real') as it. That 200 simulations achieve a result within 1% of the true probability of "at least one success" expresses what happens in the limit as more and more simulations are run. It would be easy to do more simulations (i.e. get a computer to run some more), but it would only trivially change the results, and thus the 200 simulations is perfectly adequate.

[1] Glickman, S.W., Rasiel, E.B., Hamilton, C.D., Kubataev, A., Schulman, K.A., "A Portfolio Model of Drug Development for Tuberculosis", *Science*, 3 March 2006, Vol. 311. no. 5765, pp. 1246 – 1247, DOI: 10.1126/science.1119299, www.sciencemag.org/cgi/content/full/311/5765/1246. Supporting material at: www.sciencemag.org/cgi/content/full/311/5765/1246/DC1. Up-to-date copies of this briefing note at: www.economics.ox.ac.uk/members/andrew.farlow. Feedback very gratefully received at andrew.farlow@economics.ox.ac.uk.

2) The Science paper informs us that the source of the transition probabilities was a "consultation process" with experts in tuberculosis research, and that the rates are typical for TB drug development. The issue would be how sensitive the results of the Monte Carlo exercise would be to any inaccuracies or biases in these estimates. The TB Alliance and Médecins Sans Frontières have raised concerns that these probabilities are not fully appropriate to the issue at hand. The TB Alliance (in a letter responding to the Science article) argues that the success probabilities are more typical of for-profit large pharmaceutical firms than of their public-private parnership approach. Without prejudging this, *if* this were the case, then the success probabilities used may be too low:
    a. In strictly for-profit settings, a third of attrition is due to things like the market turning out not to be large enough, or projects being killed off that are taking too long (hence, for example, eating in to patent value), whereas here this would be of less, or even of no, concern.
    b. These for-profit firms may also not take on projects that have low attrition rates in the first place (e.g. based on compounds already well-studied and out of, or close to being out of patent, etc.) if it would simply not be worth it from a profit perspective. We need to know whether there are any selection biases in the sorts of compounds that enter simulated portfolios that might then affect these attrition rates.
    c. In particular, the transition probabilities would not take into account moxifloxacin, which the TB Alliance claims has 50/50 or better chance of success. Moxifloxacin is an antibiotic that is already on the market and has a proven safety record. TB Alliance, CDC and Bayer are starting a phase III trial for its use in first-line TB treatment. The TB Alliance argues that Moxifloxacin might be available by 2010 and could reduce TB treatment time by nearly a half.
    d. ***If*** the success probabilities being used are too low, this would have raised the measured average costs of developing drugs, increased the average expected time till first new drug, and reduced the numbers of new drugs on average for any given flow of expenditure.

3) The equations used in the simulations (equations 1, 2, 3, and 4 in the background paper) are based on the likelihood of ***one or more successes at each given stage*** of development. This is not the same as the success probability of a given compound. The Science authors clearly draw attention to this important distinction in the background paper, and also indicate it in the Science paper by explicitly referring to the "likelihood that the portfolio will generate ***at least*** one successful compound." Though the authors discuss the "likelihood of bringing ***a new TB drug*** to market by 2010", and comparing and contrasting their findings with the "the costs of developing ***a new drug***" provided by the TB Alliance (a study that is not based on "at least one new drug" reasoning), the Science figures always relate to ***at least one new drug***.

4) In the illustrative example used in the background paper (page 1) the expected chance of one compound making it through all stages from phase I to becoming a

drug is 6%. Therefore, if 20 compounds start in phase I, on average the outcome would be 1.2 new drugs. If 30 compounds start in phase I, on average the outcome would be 1.8 new drugs. However, the probabilities the authors actually use in their simulations (starting from phase I, they list these as 0.30, 0.50, 0.65) generate a success probability of 9.75% for each compound entering phase I becoming a drug. Perhaps, as a minor point, the authors could have used this as their illustrative *example* – drawn from the probabilities they use in their simulations?

5) The object of interest in the Science paper is ***one or more new drugs by a given date***.
    a. This generates the shape of the purple curves ("Probability of at least one success" at any point in time) in both of the two Figures in the Science paper. Intuitively (and very colloquially) as we go forward towards years 10, 11, 12, 13, 14, the curve "probability of at least one success" starts flat, goes through a rapidly steepening section, then flattens out. This shape is because of the way the equation for the curve works. Intuitively, the ease with which "probability of at least one success" can be increased is based on the fact that the "probability of at least one success" is already low towards the left-hand side of the Figures, and already high towards the right-hand side of those Figures. So, for example, it is ever more difficult towards the right-hand side of the two Figures to raise this probability by even a small amount: Hence the flattening section of the curve towards the right-hand side.
    b. This makes one also wonder why the authors choose year 14, i.e. 2019 (bottom p1246 Science paper) and not year 12, i.e. 2017, when discussing portfolio simulations 1 and 2, but especially portfolio 1. Success by getting "at least one drug" by year 12 is pretty much the same as by year 14 in both Figures, but the extra capital costs (remembering that pharma capital costs are rising more rapidly than they are being discounted) adds to the cost measured at year 14 in both simulations (though not by much in the second simulation), but adds very little to the probability of success. On this slightly more optimistic choice of years, the authors' figures generate 5% success by 2010, but an extra seven years (and not nine) pushes this up to 73%.

6) Looking at the authors' equations (background paper):
    a. Equation 1 in a sense traces through the product development 'tree' from pre-clinical to the end product. Equation 2 starts at phase I, equation 3 starts at phase II, and equation 4 starts at phase III. Each of these equations is then used probabilistically on the data they have for each portfolio they choose to simulate, and the results then added up. They perform three portfolio simulations in all.
    b. **For the first and second portfolio simulations** (mixed portfolios over preclinical, phase I and phase II of US trials) an error on the probability of failure at phase 3, p3, (i.e. towards the end of the product development

process) is more likely to weigh in the eventual outcome *compared to the 30-compound third portfolio* simulation, than would a mistake on probability p0. Intuitively, an error in p3 will weight more heavily in the simulations of portfolios 1 and 2 than if the portfolio started 'from scratch', or more from scratch as is the case in portfolio 3. This is because the current 'up-and-running' portfolio contains compounds that have gone past the preclinical stage, and some compounds that have gone past stage I, being now at stage II.

c. For a portfolio that does not yet exist, the faults in ***any probability*** have more balanced impact (since all compounds face all the relevant probabilities during the 'adding up' process).
d. For the second portfolio simulation (to the extent it does not double up on the preclinical compounds, given that the authors are a little unclear on this, as we will see below) this extra bias *would be* more if preclinical compounds formed part of the calculation.

7) It is puzzling to see the portfolios the authors choose. The authors ask the question "What is the likelihood of bringing a new TB drug to market by 2010?" (Science paper, p1246) with the wording suggesting that the question relates to the full global TB drug portfolio. They say that there are 27 compounds in the current global TB drug pipeline, but their first portfolio simulation is based only on "clinical trials performed in the United States", which turns out to be just 11 of those 27 compounds (4 preclinical, 5 stage I, 2 stage II). On this portfolio, the authors generate a 73% probability of "at least one drug" by year 14 (we just said that year 12 was very nearly just as good). What policy advice does this generate, if the portfolio is not the *global* portfolio?

8) We are told that "over the lifetime of the portfolio" in the first portfolio simulation, the probability of getting 0 drugs is 27%, of producing 1 drug is 40%, of getting 2 drugs is 24%, of 3 drugs is 8%, of 4 drugs is 1%. Hence, the probability of getting **at least one drug** is 1 minus the probability of getting 0 drugs, i.e. 1.00-0.27 = 0.73, i.e. 73% (see Figure 1 in the Science paper). However, this 73% **category** covers all cases including ***more than*** one new drug. Indeed, just over half of this 'category' covers cases of getting only 1 drug, but the category is also composed of some 2, 3, and 4 drug scenarios (the latter being a tiny proportion of the total). The average number of drugs produced is 1.16. So, even the first small portfolio the authors simulate will produce 'on average' just over 1 new drug – but the portfolio does run the high risk of getting none.

9) These 11 compounds are drawn from a much bigger pool of compounds:
   a. Clearly if other trials are going on elsewhere, this will push up the 'global purple curve' ("probability of at least one new drug"). Applying standard portfolio thinking (and being very intuitive about it here) one does not just average the "probability of at least one new drug" generated by the sub-portfolio the authors simulate and the "probability of at least one new drug" generated by the sub-portfolio made up of 'all the other' (in this

case 16) compounds, in order to find the overall impact of the two portfolios on "probability of at least one new drug". Adding the two parts of the global portfolio ***together first*** and then simulating is what is required.

b. The key issue is what does the portfolio being analyzed **add** to the overall portfolio? In scenarios of interest, it can add much more than its own 5% or 10%. It is all down to the way portfolios work. Again, it is therefore not clear what one can read into the 73% and 93% figures from a public policy and funding priority perspective, based on portfolios 1 and 2 if these are only part of a much bigger global portfolio. *It is actually quite misleading to separate out sub-portfolios and discuss their value in isolation.*

c. In the background paper, the authors show the two sets of costs for the US and for Uganda, but they do not add together both of these pools of trials to get the global "probability of at least one drug". That would have been interesting to see. It is not clear why they just used the limited US-only portfolio in simulations.

10) ***If*** the first portfolio was the relevant portfolio, the useful lesson here would be that the chances of getting nothing from portfolio 1 are high, at 27%, and the first portfolio is suboptimally small. 27% chance of nothing is way too risky to be running with:

a. If one was running such a portfolio and if it was the only portfolio present, and if these figures are correct, one would be very worried by the 27% chance of getting nothing. The danger is that those involved with the effort could be doing a very good job and still get no drugs, but funders would not understand what was truly going on in those states of the world where no drugs are derived (funders would read in to the outcome that the initiative had 'failed'), and turn against further funding, even though those involved have been heavily constrained not to have enough compounds in the portfolio from the start. The result would be collapse and shift back to an even more expensive drug development approach. It would be best to avoid the chances of this. The Science paper would be usefully warning about this.

b. But surely, the first portfolio is not the relevant portfolio to base this thinking upon?

11) The Science paper then says that it performed a second portfolio simulation that doubled the compounds in phase 1 and phase II trials: "We doubled the number of phase I and II compounds in the portfolio" (Science paper p1247):

a. This seems to create a portfolio therefore containing 4 preclinical as before, but now 10 phase I, and 4 phase II.

b. ***Over the lifetime of this portfolio***, the chances of getting 0, 1, 2, 3, 4, 5, 6, drugs in the Science paper come with probabilities 7%, 22%, 27%, 24%, 13%, 5%, and 2%.

c. On average ***over the lifetime of this portfolio***, ***2.37 drugs*** are generated, though this figure is not pointed out in the Science paper. When working

out the financial value of any portfolio, pharmaceutical firms and investors would be interested in the **other drugs** likely to come out of the portfolio and not just the first drug.

d. This 2.37 is high, in part because of the doubling of phase II compounds, so this is slightly exaggerating the chances of getting drugs compared with an expansion of compounds at the very start of the development process, or compared with a 'more even' expansion of the portfolio.
e. Under this second simulation, there is still a 7% chance of getting nothing. However, a useful point to have flagged is that the chances of getting nothing falls dramatically from 27% to 7% by doubling the number of candidates in phases I and phase II. That is, a doubling had the affect of quartering the chance of getting nothing in this case (again, this is the way the portfolio thinking works).
f. Confusingly, the Science text says (p1247) "we doubled the number of *phase I and II compounds* in the portfolio", and the background paper also says (including typos) "double the of *compounds number in clinical testing*." However, the diagram at the bottom of page 1247 in the Science paper says "Simulation model if the number of compounds *in preclinical and clinical testing* in 2005 is doubled". And the background paper says (p3) "We repeated this model… double the number of compounds in *preclinical and clinical* tests in 2005" Which of these simulations did the authors actually do?

12) Observe how adding these compounds created a cost bulge in Figure 2 in the Science paper (especially observe the way that the columns to the left rise heavily with no impact on probability of getting "at least one new drug"). Obviously this 'adding' would hardly do anything to raise the probabilities of success by 2010, and yet it would raise the 'average costs' of any outcomes that do happen. Then (moving left to right in the Figures) costs get quite flat, whilst the probabilities of getting "at least one new drug" shoot up. This pattern picks up the early surge in costs as compounds are fed through in an attempt to get more compounds in to stage III, even as few new drugs come out at the end of the development process in the first few years of this exercise. It is no wonder that even with this doubling, "the probability of developing a novel anti-TB compound by 2010 remained less than 5%". Adding to phase I and II will not give enough time to generate "at least one new drug"; this does not suggest that adding to the portfolio would not be a good idea or valuable in the long-run.

13) A line in the Science paper says: "Our model further suggests that, in the absence of any compounds currently in phase II trials the TB Alliance would need 30 compounds in phase I testing to be 95% confident of generating at least one successful drug":
    a. Observe that this is not directly comparable to the first or to the second portfolio simulations, since the authors have gone back to a calculation from scratch, in the sense of no phase II or higher compounds.

b. So, when the authors come up with $400m as the figure to target, this is based on this third portfolio simulation only. To give actionable advice on what to do to the current 'real world' portfolio to make it 'optimal' – in the sense of driving down the chance of "no new drugs" to 5% – this $400m would need adjusting to take account of the shape of the current portfolio and the chances of *it* generating at least one new drug.
c. Figure S1 in the background paper needs about 40 compounds in phase I to get this probability of no new drugs down to 5%. 30 compounds in *that* Figure would get this probability only down to about 13%. This seems to conflict with the wording in the text of the Science paper. This is unclear.
d. Observe how the chance of getting nothing was 27%. Then the authors doubled drugs in phase I and II (or did they?) and got this chance of getting nothing down to 7%. We have to go to 30 compounds in phase I (or do we?[2]) to get it down to 5%. Though these portfolios are not directly comparable, nevertheless observe the general principle that in the area of single-digit chances of getting nothing, each percent improvement in this "get nothing by a particular year" outcome is getting more costly to achieve.
e. The 5% chance of "getting nothing", is, one supposes, chosen on the basis that a 1 in 20 chance is statistically acceptable as a random chance of getting no new drugs by a particular date.
f. It would, nevertheless, be very interesting to see what economists call "Value at Risk" analysis, VaR, to work out the full impact of this very bad outcome. 5% chance of failure might be still too high, given the fight to generate public and Foundation funding in to R&D at later dates.

14) The Science paper line: "We estimate that a drug portfolio designed to produce a single successful compound would require a commitment of *up to* $400m." (earlier in the paper the authors say "cost *as much as* $400m"):
    a. It appears that this line should have been rephrased to include the words "to produce ***at least*** a single successful compound" (allowing 5% chance of no success). There is a big difference. The authors seem to mean this, but it has not been expressed as clearly as it could have been. Incidentally, on this and all other points the Glickman et al. authors were given a copy of this briefing note (its receipt was acknowledged) and its release was delayed for over three weeks to give time for clarifications and corrections. Unfortunately, none were received. If some of this reasoning is wrong, I apologize, but the authors also had time to correct it, and may yet do so.
    b. This $400m figure appears to be based on 30 compounds starting in phase I with no candidates in phases above phase I ("in the absence of any compounds in phase II", p1246). So, it is not necessarily applicable to the situation we currently have, which is not starting from this base line.
    c. Using the transition probabilities in the Science paper (see background material page 1) of 30% (probability of transition from phase I to phase

[2] If I have misunderstood this, I will happily apologize and stand corrected.

II), 50% (probability of transition from phase II to phase III), and 65% (probability of transition from phase III to product launch), each compound in phase I has a 9.75% probability of getting all the way through to becoming a product. Thus, 30 compounds in phase I trials will on average generate 2.925 drugs at the end.

d. Hence, on average (again, if we take these cost figures at face value, which we may not be able to do) we need to divide the $400m by 2.925 to get the average approximate cost of development of a drug starting from phase I compounds only. If we presume $400m is right (remembering the caveats above and below about transition rates, costs of trials, sequential or simultaneous trials, choice of cost data, capital costs, duration of each phase of development, age of the data, etc.) this yields $400m/2.925 = $136.75m, though this refers only to clinical trial costs and capital costs by year 12, and, of course is completely dependent on the data used here[3]. The TB Alliance figures quoted in the Science paper estimate the cost at $120m to $240m per new drug. So this $136.75m is in line with the TB Alliance figure, and indeed towards the bottom end of that range. The TB Alliance figures have been heavily debated in the literature, but going into their pros and cons here – never mind comparing or contrasting with the DiMasi et al. costs of developing a drug – would be far too absorbing. They are quoted here without judgment as to their veracity. The TB Alliance figures and the Glickman et al. figures may also not be comparable (they are after all based on different methodologies).
e. This compares very favorably with very rough 'back of the envelope' calculations (provided to the author by the TB Alliance, who reiterate that the numbers have not been validated by a formal study and are based on 2004 data) of a current global flow of funding into TB Drug development of approximately $81.5m per year (composed of about $33.5m from public sector sources and about $48m from private sector sources). Of the private sector funding, the TB Alliance provides about $15m (just over 30%). There may be additional private sector and public sector sources that have not been included in this estimation. Together, these figures suggest that we should not be panicking just yet. It suggests also that we should hold back on being too harsh in our judgments of the performance of the TB Alliance, whose contribution is just under 20% of the total, and who, according to portfolio thinking, may be making a contribution out of proportion to their actual percent of the total funding flows involved.
f. The ranges of costs in the Figures in the Science paper and in the background paper are very large (in most of the columns in the Figures, the lowest cost is half the highest cost). This suggests that this $136.75m is the top of the range (the authors say "as high as", as if they are talking about the upper limit). But the authors do not actually show any range for the $400m figure. It would be very useful to see that range.
g. So, when the authors say that "Clinical testing in this scenario could take 12 years and cost *as much as* $400million," really the key message is that

[3] I say this to prevent this figure being quoted as *my* calculation of what it costs to develop a TB drug.

the cost of driving the "chance of getting no drugs" down to 5% by year 12 is up to $400m in present value in clinical trials costs including capital costs starting with only phase I compounds, and not that the $400m refers to the cost of developing 'a drug'.

h. Neither is it correct to interpret this as saying that the appropriate figure is $400m starting from the current portfolio. It might well be that the authors made the correct adjustments when they derived the $100m current shortfall of funding, that we will discuss in a moment.
i. This $400m is based on US costs. The Science paper observes that trying to achieve lower costs by conducting trials in developing and transition countries (where most TB cases are to be found) suffers badly from the lack of infrastructure needed to conduct trials using best practice. Given the reluctance of private industry to develop such sites, the authors observe that "public support for fixed infrastructure will be essential to the development of new therapies".
j. This $400m includes capital, i.e. finance, costs, and therefore refers not just to out-of-pocket. Capital costs should be costed in, but required public funding flows depend on out-of-pocket cost flows.

15) One thing the Science paper could have explained better is that the TB Alliance and other endeavors are all about creating a ***rolling stock of new drugs***, and not just about the first drug.
    a. The interesting question then would be what is the ***correct/optimal portfolio size*** to run with?
    b. Having worked up to the appropriate portfolio size, one would then maintain the portfolio at this size by taking in a few new compounds at the start of the process every year while new drugs arrive at the other end of the process, keeping the portfolio size relatively constant.
    c. Going from scratch, 30 compounds starting in phase I clinical trials is deemed about optimal according to the Science paper's reasoning (though we just observed that the figure in the Background paper seems to be suggesting that a higher number of compounds is more appropriate), if 5% chance of failure is acceptable (though see the worries above also about the risks of even running with 5% failure). This is not saying, however, that going from the currently constituted portfolio to 30 is required.
    d. Clearly, going from zero compounds to the optimal size will generate a large initial 'cost' with little apparent outcome in terms of new drugs. However, once the optimal size of portfolio has been achieved, our interest switches to the average cost of each new drug that actually comes out of the process. For example, if the target was one new drug every two years, then once the optimal portfolio size has been reached, *according to the figures here*, it would seem that a yearly flow of *very approximately* $70m in clinical trial costs (including capital costs, and therefore this would suggest a different lower figure for out-of-pocket funding flows) should do it (an exact figure would need more precise thinking than this, such that this figure should not be quoted without all its considerable

caveats!). This is not the total cost of developing a new drug, since that would have to include preclinical costs.

e. An optimally sized portfolio may have an average eventual outcome of 2.925 drugs (see calculation below) that is uncertain (it has a 95% confidence interval), but intuitively, as the portfolio rolls over, in the limit the 'probability of no success' would approach zero.
f. It would be useful to have some figures to work some of these things out.
g. The message is that policy makers and funders need to hold their nerve.

16) The Science paper also says that TB treatment needs difficult multidrug regimes and that the goal is "development of new anti-TB therapies," (i.e. plural). It would be interesting to know how likely it is that any multiple new drugs generated by the TB portfolio will be *very different* from each other, or very similar. If they are likely to be very different, then this could be particularly valuable, but unmeasured by simple cost and funding flow figures.

17) When the authors of the Science paper say that if "the estimated global TB drug portfolio generates a successful compound in year 10", the mean net present value of the costs of the development ***of that drug*** is estimated at $98m, with range of $56m to $152m, this really refers to the range of costs to get to that point in the unraveling of the portfolio ***should it happen*** to produce a drug by year ten. This $98m also seems to refer to the US data only (see Figure 1 year 10 and the background paper diagram where the US and Uganda are side by side and this is even clearer to see). The authors explain in the Science paper that by moving all trials to Uganda, this $98m is reduced to $45m.

18) In working out costs, the authors say they used cost of capital (i.e. finance) based on DiMasi et al., 2003,[4] based on the "mean cost of capital for the pharmaceutical industry from 1994 through 2000". Cost of capital is the *expected return*[5] that investors forego *during* development when they invest in pharmaceutical R&D instead of an equally risky portfolio of financial securities. In the DiMasi et al. study, half the costs of drug development are allocated to the cost of capital, with two thirds to three quarters of pre-clinical research costs being capital costs:[6]

a. It is perfectly correct to ascribe cost of capital to both publicly-funded and privately-funded activity. The authors work on the hypothesis that it is the same rate for both. Whether this is the case or not, is for others to discuss and is not treated as an issue here.
b. However, in working out cost of capital, using the years 1994-2000 would not be a good idea; it would pick up too much of the stock market 'bubble' of the late 1990s that peaked in 2000. This would exaggerate upwards the required rate of return, and therefore any cost calculations based on it. In

[4] DiMasi, J.A., Hansen, R.W., Grabowski, H.G., "*The price of innovation: new estimates of drug development costs*", Journal of Health Economics, Vol. 22, 2003, 151-185.
[5] Italicized since we only ever have actual returns and have to read in what were investor 'expectations' of returns.
[6] This is not atypical. The OTA study in 1993 found that 65% of the cost was cost of capital.

effect, the Science authors would be extrapolating forward 15 years or so using a cost of capital figure based on the late 1990s atypical stock market performance.

c. However, DiMasi et al. do not base their cost of capital figure on the period 1994-2000 but on the period 1985-2000. Using the Capital Asset Pricing Model (CAPM) they find that real capital cost per year has risen from 9% in their previous study[7] to 11% in the 2003 study. This is what Glickman et al. seem to be using, though getting the years wrong. Where this 11% comes from is nevertheless still important to us.

d. Ideally, when making a decision as to whether or not to invest, the way to do things would be to compound each (expected) investment cost by the (expected) cost of capital operative at each point in time from the moment the investment cost was incurred (i.e. the return that investors forego *during* development, on the presumption that they could move their financial investments around to maximize their financial return[8]). CAPM is a static model, and tries to get close to this 'dynamic' notion of capital cost by averaging measures of investor return at specifically chosen dates. The relevant period in the CAPM calculation, *if it is being used to make a statement about typical expected costs of developing products*,[9] should ideally be a long historical series to avoid anomalies such as bubbles and crashes affecting this averaging process; CAPM cannot handle pricing inefficiencies, so the best way to deal with them is to lengthen the data sample and not to pick atypical years when producing an 'average' measure of return. For example, the S&P 500 Composite index long-term average real rate of return (the 'market portfolio' for a CAPM calculation), taken over 100 years or so of data is about real 5.5%, which is lower than the late 1990s experience. The more the data series used in the CAPM is narrowed down to years in the 1990s, the more the return of the market portfolio reflects stock market performance in *that* period. This is not to say that that would not be a good measure of opportunity cost *then*. The issue is whether for the purposes of a study about *expected* costs of developing TB drugs, one should extrapolate forward on the basis of stock market performance over a particular past period of stock market performance.

e. Tufts estimate the cost of capital based on the years 1985, 1990, 1995, and January 2000. Markets grew especially rapidly in the 1990s with January 2000 pretty much at the height of the bull run, some would even argue stock market 'bubble' (and hence an especially bad moment to choose for entering into the CAPM averaging process). Table 2 of the Tufts data quotes the average cost of capital for each of these key years (1985 =

[7] DiMasi, J.A., Hansen, R.W., Grabowski, H.G., Lasagna, L., "Cost of innovation in the pharmaceutical industry". Journal of Health Economics, Vol. 10, 1991, 107–142.

[8] Ignoring any transaction costs of doing this.

[9] Because we should allow for required capital costs to vary and thus produce variable average cost of drug/diagnostic/vaccine development. The issue here is how much can be read from one period's average into a totally different period, and how much we can make statements about 'the' average cost of development based on a particular period.

10.8%, 1990 = 10.6%, 1995 = 11.1%, 2000 = 11.9%,) exhibiting an upward trend. The mean of this data, 11%, is then derived and applied to the out of pocket costs of *all* periods. In other words, the later period pulls the average cost of capital up, and that average is then applied to earlier periods when capital costs at the time had been lower. The bigger any 'bubble' element, the bigger the bias. Given the greater effect of compounding on longer series, and given that many of the drugs under consideration in the study would have been in preclinical trials in the earlier periods (and we know that by the end of the whole process 2/3 of the cost of this phase will be capital costs) this means that capital costs of the preclinical phase costs risk being *retroactively* pulled up by any stock market bubble present. From a financial point of view, if investors had had access to a well-diversified set of assets in all periods then they could never have made a higher rate of risk-adjusted return in the earlier periods to these investment than the ones quoted in Tufts Table 2 (and, indeed, they could have made a greater rate of return than 11% in the later periods, but over much shorter periods).

f. The drugs in the Tufts study were first tested in humans between 1983 and 1994. If these drugs entered tests at a constant rate, the average date of first test would be about 1988/9, giving each drug an average of about 11 years running up to the year 2000. On average therefore these drugs were entering the compounding process at the time when financial markets were coming out of a major crash, just before entering the long bull run of the 1990s. The threefold increase in estimated costs of their development in just one decade as quoted in Tufts could be partly because of higher clinical trial costs, partly the result of the failure to remove this bubble element, partly failure to control this timing issue, or partly because the pharmaceutical industry became genuinely more risky in this period compared to the market.
g. Since no information is given in the Tufts study regarding firm betas (i.e. a measure of their riskiness compared to the market), we do not actually have information to determine whether the pharmaceutical industry became relatively more or less risky in the period. Tufts do suggest however that the change was slight "*The real cost-of-capital in pharmaceuticals has increased since the mid-1980s* ***primarily*** *as a result of higher real rates of return required by holders of equity capital* ***during the 1990s***," suggesting that there is a bias if this data is then simply extrapolated forwards. This bias may not be huge. It may be of the order of a couple of percent or so, but, given the lengths of compounding, this can add not inconsiderably to costs.
h. DiMasi et al. have not yet updated to use, say, 1998-2003, which would have pulled this capital cost measure down by incorporating a collapse in the stock market; but this figure would have been equally problematic. If trying to calculate 'typical' costs of developing drugs, it makes more sense

to use a more long-term measure of stock market performance, and get away from the peculiarities of the 1990s.[10]

i. Even just 2% or 3% per year extra capital cost adds greatly to the total development cost because of many years of compounding. At an extra 3%, after 12 years every $1 financial capital sunk in year 0 adds 47 cents to capital costs. Percent-wise, given that preclinical costs are the same (and high) whether trials are performed in the USA or Uganda, this raises the Uganda-based costs disproportionately, when these costs are measured to include preclinical and clinical costs.

j. In DiMasi et al., 2003, the 11% is the ***real*** rate of return (i.e. after inflation is accounted for), with nominal rate of about 14%-16% (i.e. an average of about 15%). At no point does the Science paper and backup paper clarify whether the 11% rate of capital costs used is *nominal* or *real*. In fact, it seems to be interpreted as if it is nominal, even though DiMasi et al. derive it as real.

k. In the Science paper, costs incurred in the future are discounted at the "current rate on short-term risk-free instruments" of approximately 4%. It is not explained whether this is nominal or real either, but it seems to be nominal. If it were being used as a real rate (i.e. accounting for inflation too) this would make the nominal rate about 7%, which is clearly wrong. Current nominal interest rates are about 4.0% (actually 4.5%, but we can happily let off the extra 0.5%)[11]. Real rates are lower, at about 2%.

l. If so, this seems to indicate that the authors of the Science paper have used in their calculation of capital costs, real rates for pharma cost of capital (11%), but nominal rates for discounting (4%). This seems unclear. Did they in fact use 15% and 4% nominal rates? Or did they use 11% and 2% real rates? Or did they use 11% real and 4% nominal. If they had done the latter, this would have produced a lower drug development cost figure in the paper than a DiMasi et al. based costing would have done (since the authors would have been discounting costs incurred in the future at 4% instead of the 2% required to match the real DiMasi rate). Or is everything nominal and the authors, in effect, presumed that a lower real rate than DiMasi was more appropriate?

m. The authors use "The current rate on short-term, risk-free instruments" to discount future costs. It is not clear why they do not use long-rates for long-term projects. Having said that, current long-term discount rates look pretty similar to short-term rates, suggesting that this will not have biased the cost much. This is not always the case however, and it would probably be better to use a rate on long-term, risk-free instruments. This would also prevent short-term temporary fluctuations from making measurements

[10] It may be that it is difficult to use long data series given important changes in the pharmaceutical industry over the last 20 plus years. There is an inherent problem here; it may be easier to average over shorter time periods but less statistically meaningful; it may be more difficult to average over longer periods but potentially more meaningful. At the very least, the problems should be eluded to when short data runs are used.

[11] www.federalreserve.gov/releases/H15/Current

unnecessarily volatile. Even long-rates are not entirely without their problems since they are also a function of global financial balances (hence imbalances), but at least they are a better measure of market expectations which is the issue for long-term projects such as long-gestation R&D.

n. Later we will explain how we need to take care how we interpret "required returns to capital" once PPPs, Foundations, and government sponsors get involved, and what this means for 'overall' levels for required out-of-pocket funding.

19) Inputs to calculations of trial costs include the minimum and maximum costs of trials at each stage. The Monte Carlo simulation does random drawings over this cost space, and the probability distribution over this cost space is uniform (i.e. flat). In the Science paper, therefore, calculation of overall cost is based on costs that themselves depend on the assumed limits of this cost space; random choices are made over this space in Monte Carlo simulations, but this cost space is not itself random. Similarly, costs are based on a particular assumed probability distribution over this space. Again, this may not be a problem, but it does mean that the figures depend on the *assumptions* made about costs and distributions over costs and must not be treated as any more 'scientific' than that. Given the number of simulations (200), approximating a normal probability distribution over this cost space with a uniform probability distribution is not an issue. The issue is whether there is any skew in the original normal distribution, leading in one direction to under-reporting of costs, or in the other to over-reporting. The presumption is that there is no skew (unlike in the DiMasi et al., 2003, cost data for example).

20) The authors calculate the present value of "expected costs of successful and unsuccessful compounds" and presume the costs in each case are the same. However, on average, unsuccessful products tend to absorb lower costs; certainly this is the case for commercial products, where failing products tend to be dropped earlier on average, and there is also good evidence that less is spent on trials for products 'less likely' to succeed (smaller trial sizes, etc.). However, if the unsuccessful are treated as costing *the same* as the successful, this biases the overall cost of development upwards. Incidentally, something similar happens in DiMasi et al., 2003, since the costs of non-abandoned compounds that obtained marketing approval are used to estimate the costs of abandoned compounds. There is also a significant difference in the Tufts figures between mean and median trial numbers, and hence mean and median trial costs. The distribution is skewed (Table 1, page 162), with the mean cost higher than the median cost. In all cost calculations however the Tufts paper uses the mean figure. If it happens to be that trials that fail are nearer the median size, this will bias up the measured cost of trial failures, and hence the compounded overall cost. Intuitively, one is using a higher average cost for failed trials than the trials did on average actually cost. The message is that we would need more cost data on the unsuccessful products. We are told in the Science paper that these are TB-related cost data, so this may

be different from typical for-profit industry figures, since, one would presume, these particular biases are less likely to be present.

21) Similarly, the authors make random drawings uniformly over a range of durations for each development phase. Again, is the data appropriate for the analysis?

22) The methodology presumes sequential trials. However, the TB Alliance has a strategy to run simultaneous trials to cut this time down. Indeed, given the much lower costs in Uganda, yet high capital costs, there is strong logic favoring this approach. To shorten times, TB Alliance is working with regulators to see if trials can be run with two investigational drugs together. To the extant that these efforts are successful, this would cut the height of the cost columns (because of less capital cost compounding and maybe through economising on other trial costs), and pull the purple curve ("probability of at least one drug") to the left and make it steeper 'earlier on' (speaking very colloquially here). The goal of the TB Alliance is to come up with a completely new treatment regime (currently four drugs), by replacing one drug after the other. This would be much quicker via simultaneous trials and trials involving several compounds together.

23) The fall in average costs for the 13th year is a little puzzling (especially noticeable in the background paper Figure, but also noticeable in the two Figures in the Science paper).
    a. This might suggest a bunch of compounds reaching the end of a trial phase and dropping out, and that this then somehow shows up in a discontinuity in the cost figures. This does not seem very random. Besides, it is likely the methodology was done right, and that this just looks a bit odd.
    b. One suspects that the average cost of getting "at least one drug" (not worked out here) would actually fall in year 13 compared to year 12 and year 14 (since all three years are running close to either 73% or 93% "probability of at least one drug"), which would be a little anomalous.
    c. One can only think that this is maybe caused by the underlying distribution over trial phases interacting with the way the costs over the phases are compounded and added up. Since portfolio simulations 1 and 2 pick up a portfolio on-the-run, rather than starting from scratch, this affect comes out of the calculations. This is just a guess, however.
    d. At least this indicates the dangers of picking particular years to calculate the average cost of getting "at least one drug". One should take care that an analyst or a journalist, by a simple change of just one or two years in the horizon, does not lower or raise this average cost.
    e. One would expect that the 30-compound portfolio (all starting from phase I and all calculated starting from phase I) would trace out a 'smoother' set of cost columns.

24) The confidence intervals around the net cost at each year are large. It would have been interesting to see the *shape* of these intervals.

25) Average costs are also based on a 100% certain rate of capital costs. If this capital cost was measured as much more uncertain, this would increase the uncertainty of the overall cost figure, and the confidence intervals would widen.

26) The 95% confidence intervals on costs show that more of the confidence interval is above the average of these costs than below it. This is more the case on the early years, and less the case on the later years (see Figure S2 in the background paper for example). This suggests some sort of skew in the underlying data. But the probabilities over cost space and duration space are uniform (i.e. flat) and must also be presuming symmetric normal probability distributions. One suspects that this skew is caused by the underlying distribution over phases interacting with the way costs are compounded (i.e. the skew is not coming from the probability distributions but from the adding up). As soon as compounds fail, they exit the compounding process, while those that do not fail keep going in the compounding process. It would be interesting to see the pattern of confidence intervals for the 30-compound simulation. Again, one would expect the 30-compound portfolio not to behave so much in this way.

27) The TB Alliance data used in the paper is 6 yrs old (it is based on the 2001 TB Alliance Report, "The economics of TB drug development" based on even older data). Have typical trial costs risen or fallen since? Are they expected to rise or fall in the future? What has happened or is expected to happen as clinical trials expand into lower-cost trial settings?

28) If the numbers are correct, the Science article is fair to warn that little can be done to *boost* the chances of more new drugs by 2010 compared to what the chances currently are.
    a. However, the argument that there is currently only a 5% chance of getting "at least one new drug" is a separate issue – that needs to be discussed.
    b. This is a warning in general not to set unrealistic targets (c.f. the current notion of a malaria vaccine by 2015 based on still limited numbers of trials in the face of the huge scientific challenges).
    c. The Science paper does not stimulate the currently extant global portfolio in order to derive the global likelihood.
    d. It also suggests (though this is not spelled out much in the Science paper) that neither should the outcome of "no drugs" by 2010 be treated as a failure should it happen, since the early build-up to a sufficiently-sized portfolio *was always going to be costly with little outcome*. The Science paper suggests (again, if the figures are treated as correct) that past TB drug R&D funding has been too tight and may be still.
    e. The TB Alliance and others should not avoid calculating and facing up to the probability of no new drugs by 2010 – indeed, the probability of no new drugs by any date. If at horizons of interest the probability is higher than deemed acceptable, the portfolio would need to be expanded.

29) These costs do not include discovery costs. Is this an issue when considering sourcing funds for the TB Alliance and others? Maybe not? For the TB Alliance and others the key issue is funding for clinical trials.

30) The authors explain that moving all activities to Uganda halves the costs. The authors quote this observation in the Science paper, but maybe they did not have room for the diagram taken from the backup paper that very graphically illustrates this point, which is a shame. Indeed, if one separated out preclinical from clinical costs and looked at clinical costs alone, the savings are even more apparent. Reviewing Table S1 in the Backup paper, preclinical costs are the same whether the trials are done in the US or Uganda (since preclinical costs are incurred in the higher cost setting in both cases). The transition probability from preclinical to phase 1 is 0.1, and naturally much smaller than any of the other transition probabilities. Thus, roughly speaking, preclinical costs have to be multiplied by ten and compounded at, real, 11% for just under 11 years on average. Proportionally this weighs more heavily in the overall Uganda cost figure when this figure includes preclinical costs. It would have been interesting to review what would have happened had the increase in only Phase 1 compounds happened in a portfolio involving Uganda (and perhaps even *only* Uganda) and perhaps involving some of the simultaneous trials the TB Alliance indicates that it is working to encourage. One would imagine the results would be encouraging.

31) Do scale economies rise if more compounds are run at each stage? It must be so in some cases, for example if there is better coordination and more efficient use of trials infrastructure, since this would drive average costs lower.

32) Many costs should diminish over time, since there would be a big fixed cost at the start of some of these activities (remember, the figures in the paper presume costs are constant). This would be very well worth exploring. For example, there must be large fixed costs expanding trials infrastructure in Uganda (and elsewhere), but this gets spread over more drug candidates if the portfolio is larger. Similarly, setting up or expanding a PPP has fixed costs, such that the average cost of generating new drugs via an expanded PPP fall over time as the portfolio expands.

33) The Science paper presumes that clinical trial success probabilities (and failures) are independent. No evidence has been provided to this reviewer either one way or the other on this one. It would be interesting to confirm the exact situation.

34) Clearly the TB Alliance is falling short ***if*** it is not running a sufficiently-sized portfolio.
    a. However, the Science paper has not really demonstrated how far short, if at all, including the key question as to how the TB Alliance portfolio fits in to the global TB drug portfolio.
    b. The Science paper says that the estimated shortfall is $100m on the basis of an estimated $36m in cumulative funding through 2007. It is not clear

how adjustment for inflation, etc. has been dealt with in the calculation of this extra $100m figure, though one presumes it has been done.

c. The $36m (and $100m) figure is not directly comparable to some of the $400m figure since this $400m includes compounding for capital costs of pharmaceutical companies, and therefore refers to more than just strict out-of-pocket trial costs. It would be more interesting from a policy perspective to see how that $400m translates into straight out-of-pocket costs (which the Science paper may have done but has not detailed). Intuitively, the TB Alliance and others may invest at the same capital costs as pharma firms (there is a debate about whether the 'opportunity cost' to them is the same as to 'big pharma', but let us presume that it is for now, as the Science paper presumes) but they take **their** returns in part in terms of lives saved, suffering avoided, etc. and not in terms of a 'monetary' return to their invested capital. So, it is not possible to read directly into figures like this $400m (and the $70m figure quoted above) what level or flow of ***out-of-pocket*** funding is actually needed.
d. The $36m is cumulative funding for out-of pocket costs. The authors may have been thinking of the needed cash flows for *out-of-pocket* costs, and have adjusting to come up with $100m as the "shortfall for the first new drug" in this flow of out-of-pocket costs. If so, it would be very helpful to see the method of derivation of this $100m cash flow requirement.
e. Again, observe that this is not relative to the costs of developing new drugs per se but relative to the expenditure 'needed' to generate "at least one new drug" with an 'acceptable' probability of getting no new drugs from the portfolio (here chosen as 5%).
f. The authors do not show where the $100m figure came from, but the authors would have had to have made a lot of adjustments to adapt to the shape of the current portfolio and the probabilities of success of the current portfolio, to adjust for capital costs, etc. in order to produce the $100m figure. How did they do this? The $100m figure is actually one of the more important figures in the paper, since it is where the authors offer specific policy advice. It would therefore have been useful to have seen laid bare the methodology for deriving the $100m figure.

35) One key message of the Science paper is that it would be a real shame if the various global TB initiatives lost out to those lobbying for multiple ***billion dollar*** alternatives if this failure is simply because the current portfolio was starved of sufficient funds to be large enough to avoid the bad outcome of no drug.

Perhaps the authors of the Science paper need to be a little more aware that they should be careful not to leave too many unguarded comments and unclear loose ends that might be picked up and interpreted, or used, in ways that were not intended. Maybe some of these notes will help clarify what is actually going on in this study and also in TB drug development in general and help others to be prepared should an unguarded comment take on more weight than the Science paper gave it. There are also repeated bits of imprecise explanation in the Science paper and background paper that at times make it a

little unclear what the authors have actually done or are trying to say. Furthermore, the derivation of some important policy-relevant figures is not spelled out.

Any warnings the Science paper makes that are relevant to the TB Alliance and others working on TB drugs, and that are supported by the evidence, should be fully taken on board and acted upon. At the same time, the objective of the exercise is not to be seen to be doing something just for the sake of it, but to be feeding good compounds in to the global TB portfolio. This requires improvements in, for example, translational research so as to increase the chances of having a good 'rolling portfolio' of TB drug candidates, and improved and expanded trial infrastructure in low-cost settings to help rapidly feed those compounds through trials with as much room for simultaneous trials as possible. This is the best way to increase the chances of getting new TB drugs that are affordable to those who need them the most, and that are also a cost-effective use of limited R&D resources.