

# 1 ESTIMATING THE EFFECTIVENESS OF DCP-

## 2 - Post 2

Warning: I'm now going to start assuming some serious statistics knowledge! Not everyone will be able to follow the guts of the modelling, but the bottom line should still make sense. As before, we're using R to do most of the work, and the script is attached.

In the previous post, I explained a simple statistical model for assessing the effectiveness of donating to DCP. In this post I would like to describe some improvements to the model. The most important of these is allowing for the possibility of error in DCP's assessment process. Some of the variation in the DCPs results are due not to real differences in the cost effectiveness of treatments, but rather simply measurement error on their part. We will run simulations where we vary how large this error is, and see how much it reduces the value of DCPs work. The conclusion is that given our guesses about the size of the errors and the money their research moves, DCP still appears much better to fund than the best treatment they identify. Firstly, however, we can be a bit more sophisticated when it comes to estimating the parameters of the distribution of interventions. In the previous post, we simply calculated the mean and standard deviation of our data, and assumed that these were a good approximation to the true values. In this case, they probably are a decent approximation, but we can do better using Markov Chain Monte Carlo methods<sup>1</sup>. Let's set up some of the formal machinery.

Throughout this post we're almost always going to work with the logarithms of our data. So, unless otherwise stated, you may assume that that's the case. Now, let's suppose that the true cost-effectiveness of the  $i$ -th intervention,  $t_i$  is normally distributed with mean  $\mu$  and standard deviation  $\sigma$ ;  $t_i \sim N(\mu, \sigma^2)$ .

We have a lot of data taken from this distribution (that is, the DCP data). Call this  $D$ . Given two values for  $(\mu, \sigma)$ , we can work out how likely it is that we would get the data that we have got, given those two parameters, i.e. to abuse notation somewhat, we can calculate  $P(\mu, \sigma | D)$ . This lets us codify our uncertainty about  $(\mu, \sigma)$  by giving us a probability distribution for the parameters themselves. To work with this statistically, we can use a Markov Chain to produce a set of samples from this distribution. We can then run the rest of the process for each of these sets of parameters, and average the results.

This will already give us a more accurate estimate than we had before, but it also gives us the tools to deal with the possibility of error. Let's expand our model further, and suppose that the observed values we have (the data  $D$ ) are taken from the distribution previously described, plus some normally distributed error. So  $t_i \sim N(\mu, \sigma^2)$  and  $m_i \sim N(t_i, \sigma'^2)$  where  $\sigma'$  is the standard deviation of the error,  $m_i$  is the "measured" value, and  $t_i$  is the true value as before.

Now our data,  $D$ , consists of a series of values for the  $m_i$ s, and we want to know the  $t_i$ s and the true values of the parameters (including the variance of the error).

However, we can extend our MCMC sampling to include the data itself, since for any sample  $S$  consisting of a set of data and values for  $(\mu, \sigma, \sigma')$ , we can calculate the probability of a new sample  $T$  given  $S$ : the new set of data will be constrained by the previous parameters, and the new parameters will be constrained by the previous data.

This is enough to run the Markov process to get a set of samples for both the parameters and the data. We do, however, need a starting point. The mean of the  $(t_i)$ s we can assume to be the mean of  $D$ , but the variance is slightly trickier. We're going to assume that the

---

<sup>1</sup>see wikipedia for details. We're using Gibbs sampling, which is a bit more complicated, but Metropolis-Hastings is easier to understand and functionally the same.

variance we measure from  $D$  is due partly to actual variance in the  $t_i$ s and partly due to error, in some proportion  $\rho$ . That is,  $Var(D) = \rho\sigma^2 + (1 - \rho)\sigma'^2$ . For the moment, let's not worry about the actual value of  $\rho$ , but just assume that we have one. Since we can measure  $Var(D)$ , this gives us a starting point for the MCMC process.

At this point we're going to do something very similar to what we did last time: for each of our samples, we're going to generate lots of values for  $t_i$  and  $m_i$  given the values of  $(\mu, \sigma, \sigma')$  from that sample. Then for each of them, we're going to try to model what would happen, according to our model, if DCP were to assess a new intervention with that value. The tricky part here is that although we know the \*true\* efficiency of the intervention, our hypothetical DCP assessor does not. All they would know is the \*measured\* efficiency. So, for each of them we can do the following<sup>2</sup>:

1. Which intervention is currently believed to be best? (i.e. which  $m_i$  is biggest in  $D$ ?)  
Suppose it is the  $j$ -th intervention.
2. Is the newly generated intervention believed to be better than that? (i.e. is  $m_i > m_j$ ?)
3. If so, assume that the money moved switches to the  $j$ -th intervention from the  $i$ -th.
4. Then calculate the difference in the true effectivenesses of the  $j$ -th and  $i$ -th interventions. (i.e.  $t_j - t_i$ )
5. Average the gains in effectiveness.

Remember  $\rho$ ? We're not actually sure what the right value for  $\rho$  is, but what we can do to start with is to plot how our estimate varies with  $\rho$ . In Figure 1 the red line is our estimate for the (non-logarithm'd) effectiveness of donating to DCP, and the x-axis is what we've been calling  $\rho$ . The green line shows the average true effectiveness of the intervention that appears to be best (when we add in the error), that is, how good it is to donate to DCP's most recommended intervention, rather than DCP itself. Unsurprisingly, as the amount of variance due to error goes to 100%, not only does donating to DCP become worthless, but our estimation of how good their current recommendations are drops to almost nothing. However, until the amount of error becomes excessive, it is still significantly better to donate to DCP than to their most recommended intervention.

Figure 2 represents our uncertainty about  $\rho$ . We aren't totally uncertain about it; it's pretty unlikely that the error is zero, or 100%. We can describe this by giving a prior for  $\rho$ . Here we've chosen to use the beta-distribution that you can see on the graph; mostly because it has roughly the right shape. Now we can incorporate this uncertainty about  $\rho$  by simply multiplying our estimate curve by our prior for  $\rho$ , and then averaging.

Finally, we still need to think more carefully about our numbers for the "money moved" and for the cost of DCP doing an investigation. The latter we can get a fairly good estimate for just by considering DCP's budget. DCP3 has a budget of  $\sim$  \$30 million, and will probably be investigating  $\sim$  300 interventions, giving the cost-per-investigation figure of \$100,000 that we used before.

Estimating the money moved is trickier. Here our current best option is to get some expert estimates. The figures we've found suggest that the median amount is probably about \$200m per annum, with the 5th percentile at \$20mpa, and the 95th percentile at \$2000mpa<sup>3</sup>. This suggests that we should expect the money moved to be log-normally

---

<sup>2</sup>In the script this looks a bit different as we're working with logarithms, but the idea is the same.

<sup>3</sup>Toby Ord, personal communication.

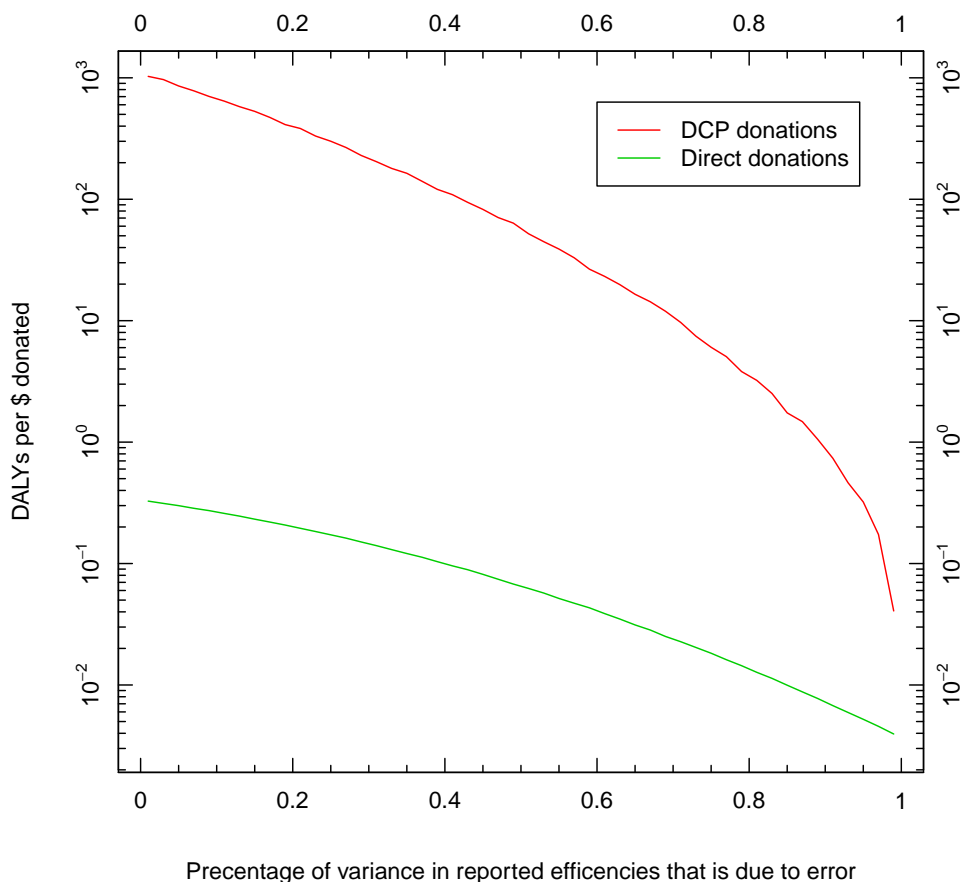


Figure 1: Comparison of cost effectiveness of donating to DCP (red) vs donating to the charity that DCP currently reports as the most cost effective (green).

distributed with these features, which has a mean of  $\sim 530$ mpa. To convert this yearly value into a lump sum we multiply by a factor of ten as intervention efficiencies are unlikely to change greatly within this time-frame.

This gives us an overall estimate of 300 DALYs/\$ for donating to DCP. This is considerably lower than our previous estimate (1000 DALYs/\$), but our estimate of the best intervention recommended by DCP has also dropped, to 0.15 DALYs/\$ (it was previously 0.33 DALYs/\$). This is a similar phenomenon to regression to the mean: if we believe that there is some degree of error in our measurements<sup>4</sup>. So donating to DCP still looks  $\sim 2000$  times better than funding their most recommended intervention!

What if we've already surveyed almost all the interventions there are? This just means that we're unlikely to actually find an amazing intervention: but the \*expected\* value is still excellent. We should expect to need to survey about 70 interventions before we find one that's better than the current best; as mentioned before, it's a long shot, but with a good payoff.

Some issues that could trip up this conclusion are the following:

<sup>4</sup>in the case of regression to the mean it's more like "performance error"), then our estimate of the actual value should be closer to the mean.

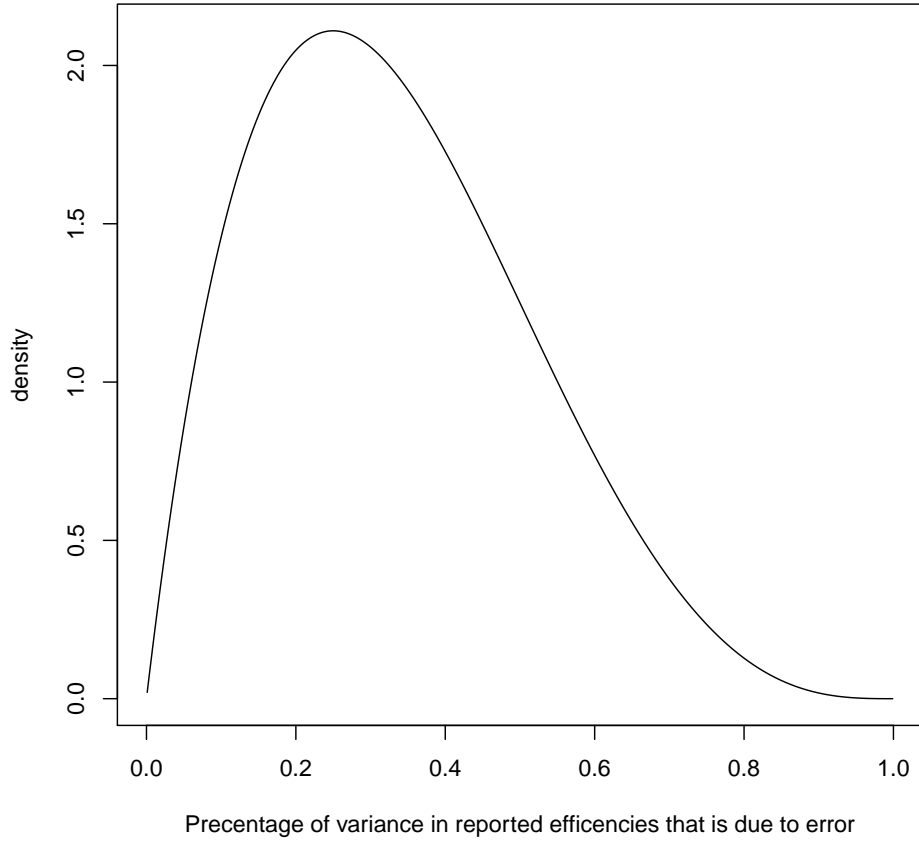


Figure 2: Prior distribution for the  $\rho$ , the percentage of variance in reported efficiencies that is due to error.

- The money moved could be much less than we thought, for example if all plausibly best treatments are already being funded
- The errors in the estimates could be much larger than we thought
- The distribution of new treatments assessed could be different from those already assessed.
- The log-normal distribution may be truncated at high efficiencies (the low hanging fruit may have already have been taken).
- A new intervention judged to be highly cost effective may not be performed by any existing charity.

Nonetheless, a 2000-fold difference provides a large margin for error from which to begin correcting for these issues