

The dangers of single metric accounting in public policy

Johanna Thoma (LSE) questions whether policy-makers should rely on a single metric for measuring the impact of far-reaching policies, such as lockdowns.

When policy interventions have far-reaching consequences, it is important to take proper account of all their relevant effects. This applies as much to decisions made at times of crisis as those taken in ordinary times. The LSE's Paul Dolan laments that, in the UK at least, there has been too much of a focus on a single metric during the pandemic — namely the number of lives lost in the short term. Too narrow a focus on this single measure has made us lose sight of costly ripple effects on long-term health (including mental health) and the economic prospects of young generations in particular.

Dolan defends a particular form that policy evaluation should take. Interestingly, a single metric also features as part of his proposed strategy for making sure everything is taken into account: all costs and benefits, short- and long-term, should be aggregated using a single metric. Policies should then be evaluated in terms of the value and distribution of that metric. The metric Dolan proposes is subjective wellbeing adjusted life years, or [WELLBYs](#). That is, what we should be aggregating is how many life years are gained and lost as a consequence of the policy intervention, where life years are weighted by a measure of the self-reported quality of people's experiences. This WELLBY metric is then used in an equity-weighted cost-benefit or cost-effectiveness analysis: gains and losses in WELLBYs should count for more when they happen to the worst-off.

WELLBYs will remind the reader of the better-known [QALY](#) (quality-adjusted life year) metric, widely used in policy evaluation in public health. But QALYs are restricted to the evaluation of health-related outcomes, whereas WELLBYs are intended to be more comprehensive. And while the measurement of QALYs typically relies on hypothetical comparative judgements people make about different health states, the measurement of WELLBYs is supposed to be based, as much as possible, on direct reports on the lived experiences of different outcomes. What WELLBYs aim to capture is a hedonistic notion of wellbeing; a good life is made up of good (pleasurable, purposeful) experiences.

Using a hedonistic metric as a measure of costs and benefits in a standard cost-benefit analysis is a way of implementing a hedonistic form of utilitarianism, according to which the sum total of subjective wellbeing should be maximised. Equity-weighting one's cost-benefit analysis, on the other hand, implements a hedonistic form of prioritarianism, according to which the experiences of the worst-off count for more.

The philosophical debate on hedonism as a theory of wellbeing, and on utilitarianism and [prioritarianism](#), is vast. Instead of offering a critical discussion of these particular moral views and their application in public policy evaluation, I would like to offer two challenges to the idea that a single metric should be used in order to aggregate the many effects of public policies and to articulate the trade-offs involved in making a decision.

Ultimately, we need multiple metrics as an input into public deliberation

The first is to point out how the use of a single metric limits the extent to which we can take 'distributional concerns' into account, making the moral commitments embodied in such approaches narrower than they might initially seem. The possibility of equity-weighting is often presented as a way to flexibly accommodate those with more egalitarian leanings. But it can only accommodate concerns about the distribution of the single metric that is being equity-weighted. And those with egalitarian leanings often also care about the distribution of other things (or indeed equality of a non-distributional kind).

The second challenge starts from the observation that the choice of a single metric and its implementation in the aggregation of diverse effects necessarily involves settling many moral questions that reasonable people (including philosophers) disagree on. This raises serious questions as to what role such a method of policy evaluation can and should play in informing policy-making in liberal democracies. Ultimately, we need multiple metrics as an input into public deliberation.

What else might matter

The hedonistic theory of individual wellbeing is only one of several theories about what is good for individuals. Some have argued that what is good for you is to get the things you desire, or would desire under some ideal conditions, whether you desire pleasurable experiences or not. The pandemic and the policy responses to it around the world have massively frustrated people's life plans. Others have argued that there are some things that are good for people whether they enjoy or desire them or not. If education, engagement with the arts, regular gatherings with family, or direct contact with different ways of life are such things, again the pandemic and policy responses have undermined these in ways that a hedonistic wellbeing measure might not capture.

Using a metric for policy evaluation based on any one of the available theories of wellbeing means leaving out or underplaying some of the things at least some reasonable people take to be morally relevant. Moreover, many people believe that there are things other than wellbeing that matter, such as freedom, or the preservation of the environment for its own sake. In this respect, too, equity-weighted cost-benefit analysis using a broadly hedonistic wellbeing measure cannot capture everything deemed to be morally relevant by at least some parts of the population. Even those who agree that the hedonistic theory is the correct theory of individual wellbeing, and that all other matters are simply 'distributional concerns', may not be satisfied that equity-weighted subjective wellbeing-based cost-benefit analysis captures everything that is morally relevant. This is because the framework cannot accommodate all distributional concerns.

As [Dolan puts the idea](#) of equity-weighting, the claims individuals have to resources depend both on the gain in WELLBYs they can expect as a result of those resources, and on their current and expected lifetime suffering or wellbeing compared to others (as well as the WELLBY effects they have on others).

But those who defend some form of distributional equality or priority often care about the distributions of things other than wellbeing. For instance, resources, capabilities, or opportunities for wellbeing are alternative potential metrics of equality or priority. Again, it is easy to think of ways in which their distributions may have been affected in ways not perfectly correlated with wellbeing itself during the pandemic. The pandemic and policy response have affected — and have most likely diminished — the ways in which resources can be translated into wellbeing, but despite that, you might think it still matters that some simply have more than others. And the pandemic response has taken away many opportunities for welfare which you might think were important for people to have, whether they used them or not.



Photo: [americo lucchi](#) via a [CC BY SA 2.0 licence](#)

So WELLBY-based analysis offers narrower ways of accommodating distributional concerns than it might seem. In large part, this is down to the ambition of using a single metric to both capture expected harms and benefits, and to capture distributional concerns.

Finally, there remains an important ambiguity about risk, or when policies impose *probabilities* of harms and benefits on people, rather than certainties. In such contexts, is what matters the *ex ante* distribution of risks of harm and chances of benefit, or is it the *ex post* distribution of harms and benefits?

Distributing harms, distributing risks

Consider the following choice problem loosely based on an [example by Peter Diamond](#). You are in charge of making sure one of your equally well-off flatmates — Amal or Bella — moves out, and this outcome would be equally bad for each. Do you:

- Choose Amal,
- Choose Bella, or
- Toss a coin?

There seems to be an intuitive equity case for tossing the coin, even though *ex post*, the outcomes of all three choices have the same wellbeing distribution. If you agree, this is likely because you think the distribution of chances of harms and benefits matters. To show how such an intuition might extend to policy, suppose that you have to choose between the following two prospects for a population of 10 million:

- Everybody faces an additional 0.002% risk each of a loss of 30 WELLBYs.
- One thousand people (at welfare levels that are representative of the population at large) face an additional 10% chance each of a loss of 30 WELLBYs.

Here, there seems to be an intuitive equity case for the former, even though it is virtually certain that the loss of WELLBYs is larger and no more equally distributed *ex post*. If you agree, then again this seems to express concern for the distribution of risks of harm, which are much more concentrated on a few individuals in the second choice.

Cost-benefit analysis in the social welfare function tradition can implement equity-weighting in two main ways in the context of risk. It can either introduce equity weights on the *ex ante* expectations of the harms and benefits a proposed policy imposes on individuals. Or it can equity-weight the *ex post* distributions of harms and benefits in the population for each potential policy outcome, and recommend the option with the best expectation of equity-weighted outcomes. The first strategy is sensitive to the distribution of risks, would recommend tossing a coin, and, with the right parameter choices, could recommend the first option in the second case. The second strategy is insensitive to *ex ante* distributions of risks, would be indifferent between the options in the first case, and recommend the latter option in the second.

There is a lively debate about which of these two strategies is better. There are also ways to combine them. But the defence of an equity-weighted cost-benefit analysis is ambiguous on this morally important question, and talk of a single metric may obfuscate the issue. Both the *ex ante* and the *ex post* approaches (and any combination between them) use a single wellbeing metric, and then merely proceed to combine this with probabilities and equity-weights in different ways. But in so doing, the *ex ante* approach implements a distributive concern for a currency different from wellbeing – namely chances of wellbeing, or conversely, risks of harm. If we think the distribution of such chances and risks matters, so if we favour the *ex ante* or a mixed approach, then there are two things we need to keep track of: the expected wellbeing distribution in the population as a consequence of policies, and how the risks of harms and chances of benefits are distributed. If we don't, our approach to policy evaluation does not keep track of things that at least some people find morally relevant.

Is wellbeing all that matters? Whose wellbeing matters?

With COVID, the risk of death or serious adverse health outcomes is much higher for the elderly and sick. Policy responses differ in the extent to which they concentrate or spread risks of harmful outcomes, such as unemployment, within the population. If avoiding an unequal spread of these risks is a distinct policy goal, this may sometimes lead us to accept lower and no more equitably distributed expected aggregate wellbeing in the population *ex post*, as in the stylised policy example above.

The problem of reasonable disagreement

There are many other ways in which implementing a WELLBY-based analysis involves settling on specific answers to a number of contentious moral questions. What is the right theory of wellbeing? If it is hedonistic in general, which experiences count as bad, and which as good? Is wellbeing all that matters? Whose wellbeing matters? What should the equity weights be? Should they be applied *ex ante* or *ex post* (or both)? These are all questions that reasonable, thoughtful and well-informed people disagree about.

The danger, as I see it, is illustrated by this caricature. If social scientists were to simply present policy recommendations based on evaluations in terms of a single metric without further context and qualification, this would not only mask all of the contentious moral decisions that went into the construction of that metric, but would also endow the recommendation with the authority of scientific expertise, making it hard for public decision-makers to diverge from the recommendation. And that would be a threat to the liberal democratic ideal of how public decision-making in the face of reasonable disagreement should be done. Value conflicts should be resolved by democratically elected officials in a way that is [open to public scrutiny](#). Of course, policy-making needs social scientific input, and recent philosophy of science is also rich in demonstrations that social science, just like any science, [can't help but be value-laden](#). But there are clearly ways in which social scientists can make sure to help, rather than undermine, democratic decision-making. When it comes to comprehensive frameworks for policy evaluation like the one advocated by Dolan, there seem to be two main strategies for doing so.

One strategy is to work closely with the public and democratically elected officials to devolve as much as possible all important value judgements, so that the resulting recommendations would have [democratic legitimacy](#). To assuage worries about a lack of democratic legitimacy, however, there would need to be democratic input on all the contentious value assumptions, which would be a [big undertaking](#). And it is not clear to me that such an undertaking would result in anything like an equity-weighted WELLBY-based analysis. For instance, there is [some evidence](#) that many people are reluctant to trade-off especially large burdens against any number of smaller burdens, which is antithetical to this framework.

The alternative strategy involves presenting one's preferred framework of policy evaluation (or its specific applications) to policy-makers as only one of several reasonable ways of evaluating policy options. But if we care about comprehensiveness in a wider sense — that there is proper accounting of all the things that reasonable people might find morally relevant — we also have to make sure that alternative frameworks and metrics are presented to the public, to enable there to be an informed public debate amongst people with different values, to reveal whether there are options that can be endorsed from any or most moral perspectives, and ultimately to give policy-makers informed options as to which values to pursue.

From within some particular value frameworks (for instance, a hedonistic *ex post* prioritarian one), the call for a single metric of policy evaluation makes sense (assuming there is transparency about what goes into the metric) and can be a means of comprehensively aggregating everything morally relevant within that moral framework. But from the wider political perspective, where the goal should be to ensure that the outputs of policy-relevant social science enable and inform public discourse in the context of reasonable disagreement, what we need are multiple metrics and frameworks.

Acknowledgements

I thank the participants of the Beveridge 2.0 Symposium on Wellbeing at the LSE for helpful discussion, my colleagues Alex Voorhoeve and Liam Kofi Bright for valuable feedback on an earlier draft, and Paul Dolan for the opportunity to comment on his work.

This post represents the views of the author and not those of the COVID-19 blog, nor LSE. It is an edited extract from Thoma, J., 2021. [Weighing the Costs and Benefits of Public Policy: On the Dangers of Single Metric Accounting](#). LSE Public Policy Review, 2(2), p.4.