"How evaluations can make you brilliant

or

Lessons about impact from Roger Federer"
Talk given by Caroline Fiennes at the Skoll World Forum 2013

This document

This document outlines the session given by Caroline Fiennes, Director of Giving Evidence, at the Skoll World Forum in Oxford, April 2013. It comprises notes and slides: it is designed to be self-explanatory but is obviously not a verbatim transcript. Many of the ideas here are discussed further at www.giving-evidence.com and in the acclaimed book, *It Ain't What You Give, It's The Way That You Give It*.

This talk

Is going to look at how we can all improve. It's about having impact, not so much about measuring it (actually the debate is weirdly skewed towards measurement and away from achievement). It's about making great decisions in future, rather than measuring and examining what's happened in the past.

Comparisons

Let's start a long way from here. In Kenya, where diarrhoea is a major problem, one of the biggest child killers, because people have to drink polluted water. This lady here is one such. It's particularly tragic because it's so cheaply avoided.

Suppose I tell you that Programme X can prevent 300 cases of diarrhoea for $1000. Would you fund Programme X? [Several in the audience said that they would.] Yeah, Programme X sounds pretty cheap. But what if we now learn that with $1000, Programme Y would prevent 400 cases? Suddenly Programme X doesn't look so great. Perhaps there's another alternative, Programme Z, which, with $1000, can prevent 700 cases? Suddenly, Programme X looks quite rubbish.

The point is that we can only see whether Programme X is any good by comparing it with the alternatives: **understanding impact is fundamentally a comparative exercise**. Actually it's rare in the charity world to be able to measure an outcome as unambiguously as here, where it's number of cases of diarrhoea, and it's even more rare to have the unit cost for the outcome. So comparisons are difficult, but we can now see that they're essential. [Discussed further here.]

For the comparison, we need a set of reliable data points. This example isn't made up: the data are real:

Diarrhoea Incidents Avoided per $1000 Spent (Kenya)

From J-PAL (www.povertyactionlab.org)
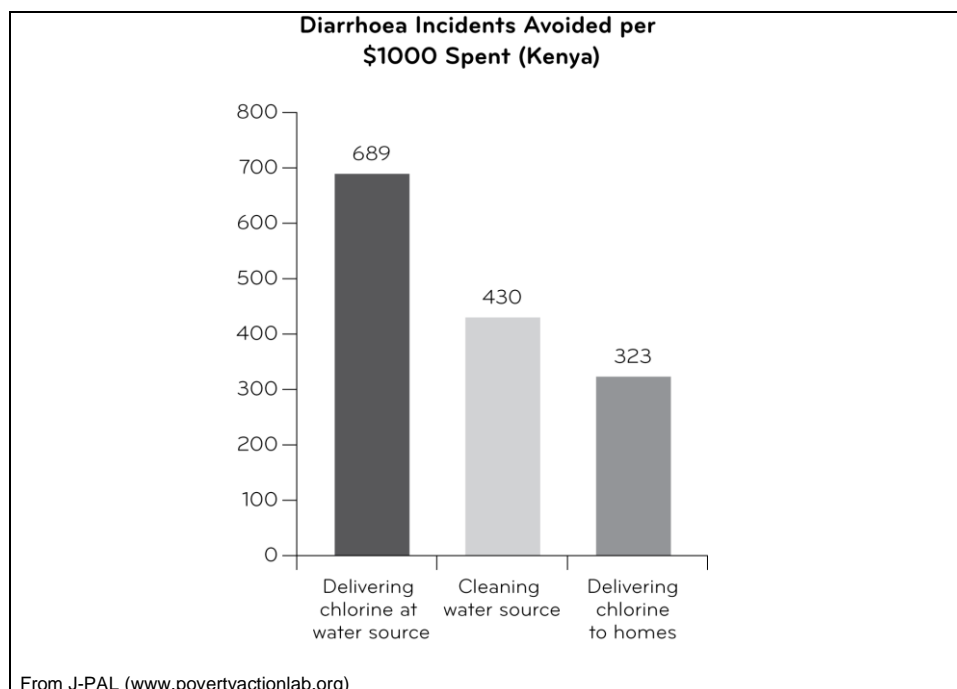
The aim of the game is finding the best. We're not just trying to achieve something but to achieve all that we can: preventing as many cases of diarrhoea as possible. Roger Federer doesn't play in order to have a nice day or win a few games: he plays to win matches, tournaments, break records. It's not about good: it's about being the best.

Economists call this 'the opportunity cost' – I prefer to call it the children who are ill but who could have been helped, or children out in the fields who could have been in school if only we'd made better decisions. It's those people, and preventing those avoidable tragedies, which motivates me in my work, and hence my book about how any donor can do a great job is dedicated to those people. Because this isn't some academic point – this is real people, like the lady we saw:

Our whole job is about avoiding opportunity costs: serving everybody who can be served. For which, we need comparisons because we're allocating finite resources between an assortment of choices. That's true if you're a funder, or an operating charity, or a social enterprise. Evaluation – or rather, using evaluations – is fundamentally a comparative exercise.

Gotta use the data, which nobody does

Bill Gates' annual letter this year was about measurement saves lives. Look how strong his language is:

> **"I have been struck again and again by how important measurement is to improving the human condition**
>
> You can achieve amazing progress if you set a clear goal and find a measure that will drive progress toward that goal – a feedback loop
>
> Without feedback from precise measurement, invention is "doomed to be rare and erratic." With it, invention becomes "commonplace."
>
> We need better measurement tools to determine which approaches work and which do not"

"...How important measurement is for improving the human condition."

We've all heard umpteen statements like this before. But it's a million miles from what's actually going on.

For instance, look at the following conclusion from the Shell Foundation which, 10 years after its inception in 2000, wanted to look at how it was doing. It looked around for benchmarks. And found basically nothing:
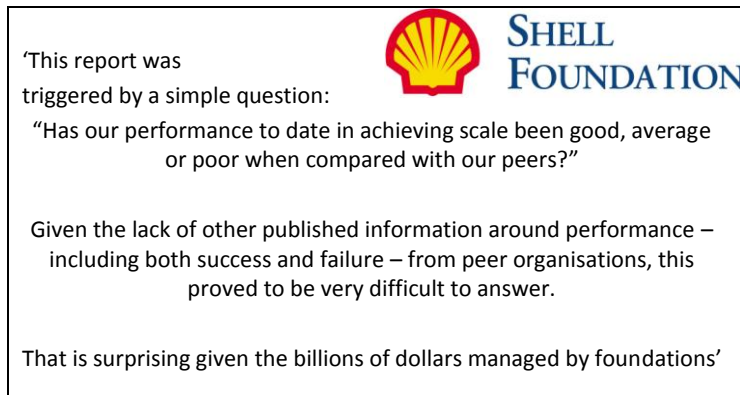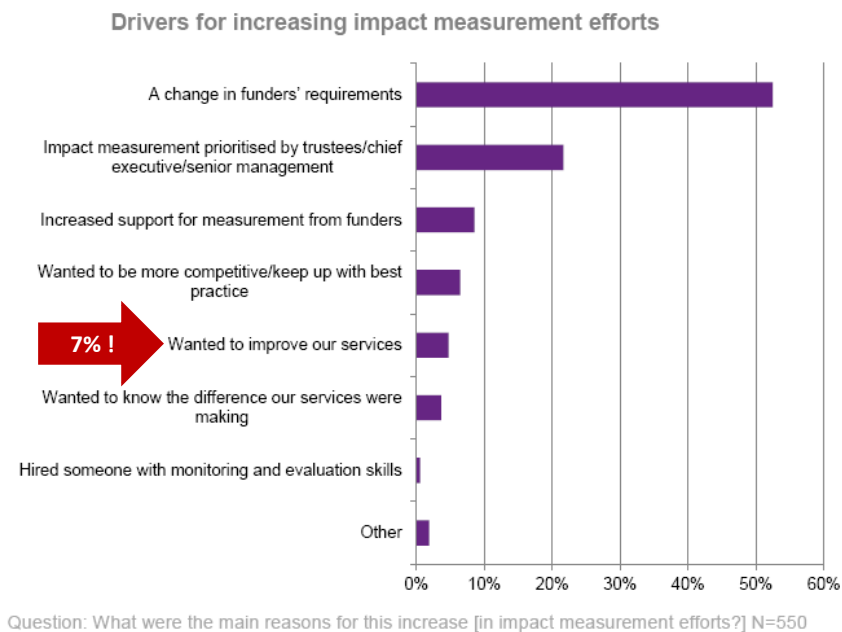
Surprising indeed. This was in a report given more than 10 years after we all started talking about impact and measurement and so on. Yet there was essentially nothing published.

Presumably, foundations don't report on their impact because they typically don't know – they've not measured it, or assessed it at all. Or maybe they have but they're not sharing. Both are absurd: how are we supposed to reduce opportunity costs – wasted lives – if we're not even trying to find out and share what we're learning?
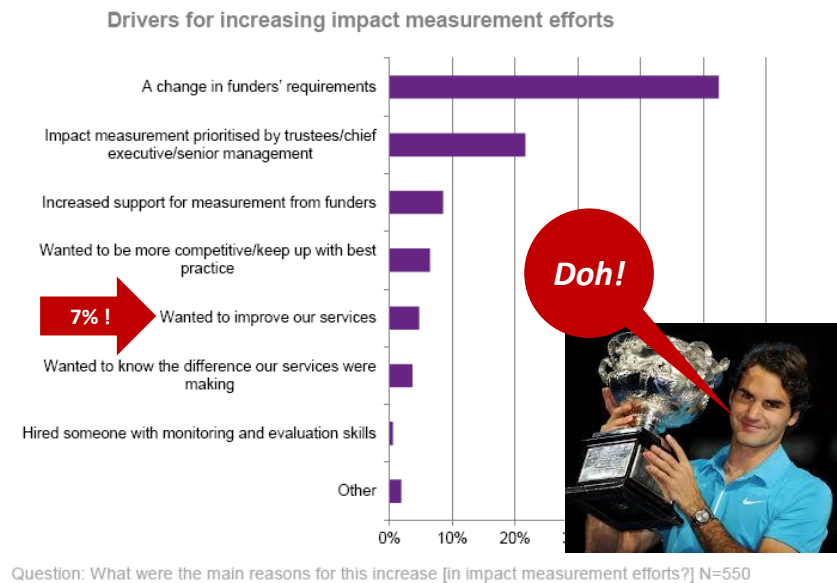
The Shell Foundation experience isn't an isolated example:

- **At a recent conference of the Social Impact Analysts Association** – people whose job involves assessing impact, obviously – you would think that people would know about these kinds of choices. About 70 of its members, who self-identify as social impact analysts, met in Berlin, and somebody asked how many had seen a programme stopped because of evidence of poor impact. Nobody. Not one.

- **UK charities and social enterprises surveyed about their impact assessment practices.** As well as asking *how* they measure, it also asked *why*. How many of them are doing this stuff in order to learn and improve? Fully seven percent! Essentially none. Rather they're producing data in order to tell somebody else:

Drivers for increasing impact measurement efforts

Question: What were the main reasons for this increase [in impact measurement efforts?] N=550

Making an Impact: Impact Measurement Across Charities and Social Enterprises in the UK, NPC, October 2012

This sounds awful – and it is awful. But it's quite rational, actually. In my own experience as a charity CEO, I found impact assessment to be totally (and stupidly) intermingled with marketing. Funders ask the charity for evidence of impact, while the charity is pitching. And in fact you don't get any more money if you're impact is good. For the most part, you measure impact in order to say something after the grant. The **impact reporting is glorified compliance** and hence not used for learning at all. That doesn't save any lives: it just costs money and time.

Let's return to Roger Federer. You don't see him monitoring his serve speed in order to tell somebody else: he does it in order to learn, to improve, to achieve more.

Drivers for increasing impact measurement efforts

Question: What were the main reasons for this increase [in impact measurement efforts?] N=550

Making an Impact: Impact Measurement Across Charities and Social Enterprises in the UK, NPC, October 2012

This phenomenon is not unique to Brits! A **survey about impact measurement practices in Ireland found** how charities and social enterprises are fearful that "impact assessment" may find them out:

> "There is a deep-rooted **fear of finding out (or 'being found out')** that one has not had the impact that was intended.
>
> Organisations are incredibly reluctant to admit that programmes have not gone to plan.
>
> Some simply do not tell funders the truth; others are very opaque when reporting back to funders; yet others cherry-pick clients to ensure low success rates are minimised.
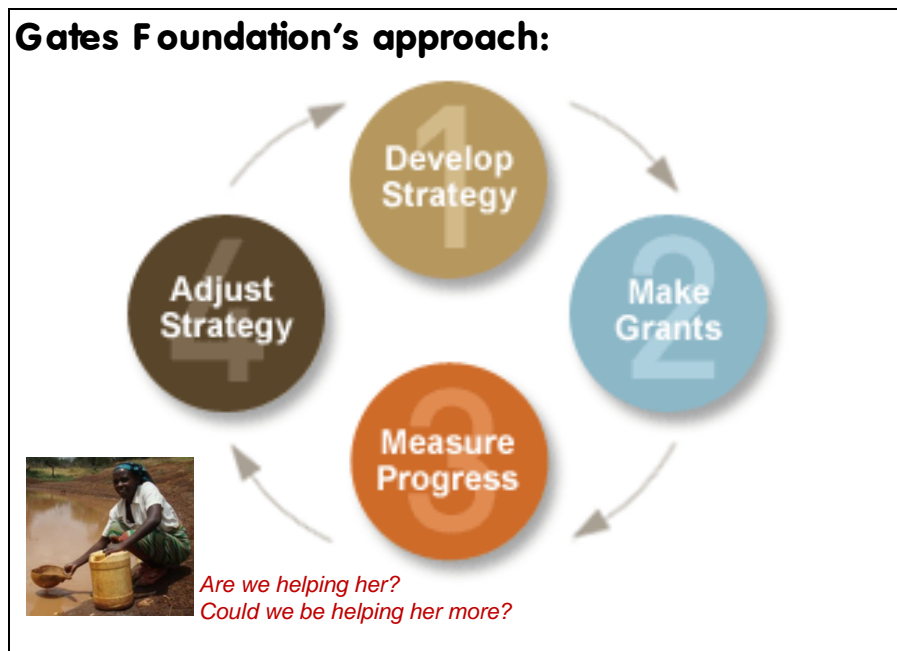>
> Lessons of 'failure' are rarely shared.
>
> When funders become aware that the desired results have not been achieved for whatever reason, they are seemingly equally reluctant to take constructive action, for fear of damaging the organisations' (and possibly their own) reputations."
>
> *- Demonstrating Impact: Current Practice Amongst Social Purpose Organisations in the Republic of Ireland*, The Wheel, 2011
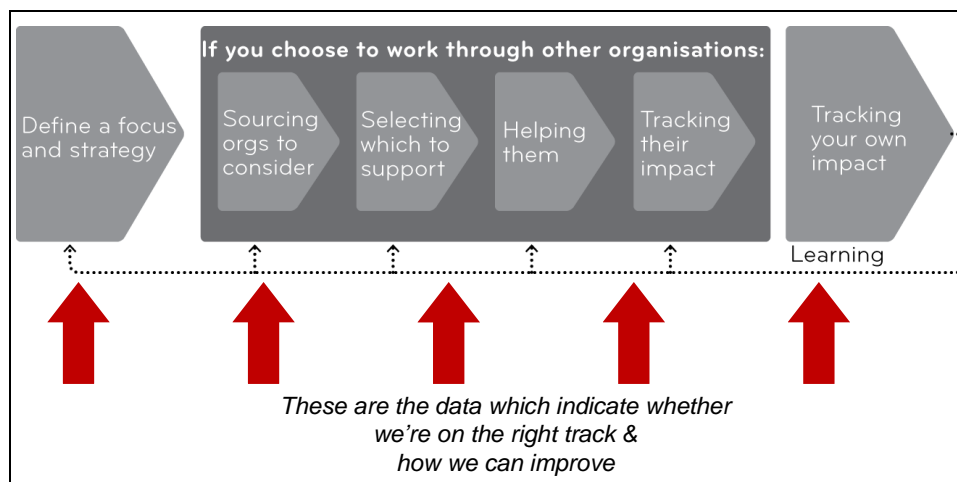
What do you notice about all of this? **Nobody's learning anything!** Nor are they changing anything. People often do measurement without looking for what it really tells them – **they're not answering any questions.**

Measurement only saves lives if you look at the answer and change your practice accordingly. Sport can teach us a lot actually because it focuses on excellence: e.g., tennis players practising their serve hit some, if it isn't working they change something until it does work. They use a learning loop, rather like this from Gates Foundation:

Gates Foundation's approach:

1. Develop Strategy
2. Make Grants
3. Measure Progress
4. Adjust Strategy

Are we helping her?
Could we be helping her more?

This would be mind-blowingly obvious for tennis. (Make a choice of sport to play; do it; see if your serve goes in the box; if not, change something until it does.)

For funders or donors, I use a slightly different flow-chart which divides up their process a bit more clearly. One reason is that in grant-making foundations, the go/no-go decision dominates their thinking, often to the exclusion of other stages such as sourcing, which can be very influential.

If you choose to work through other organisations:

Define a focus and strategy | Sourcing orgs to consider | Selecting which to support | Helping them | Tracking their impact | Tracking your own impact

Learning

These are the data which indicate whether we're on the right track & how we can improve
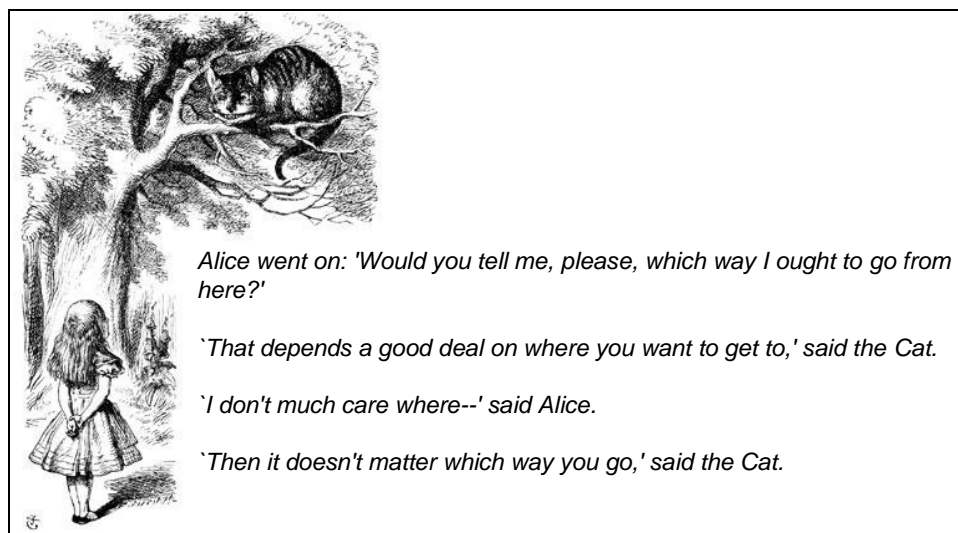
Putting this together: what to actually do

So we've got these two big ideas: that impact assessment is fundamentally comparative, and about opportunity costs. Let's combine them now.

What tracks do you want to be on? What are you trying to achieve? Charities and funders need very clear focus. Again, this sounds rather obvious, but we quite rarely see foundations with a deep understanding of the problem they're trying to solve. [There is a great example, discussed in *It Ain't What You Give*, of a foundation in Maine, US, trying to improve college results of students there. It studied extensively the reasons for the low attainment, and built an unusual programme accordingly.] Often funders have a rather 'vanilla' 'one-size-fits-all' process of being open for grant applications, selecting a few at periodic meetings, making cash-only grants.
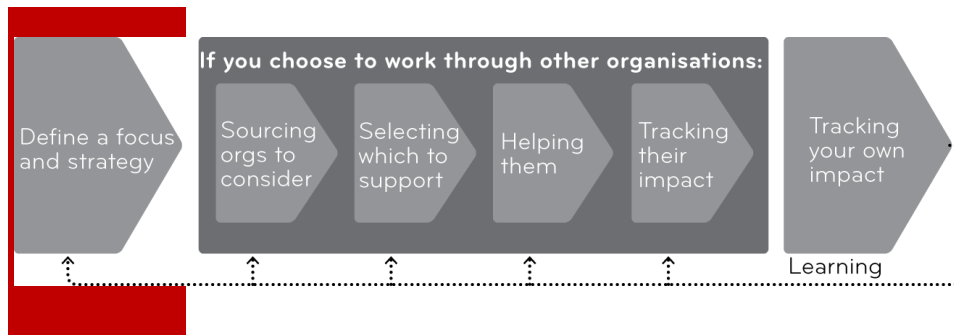
So make an extremely clear choice about what you want to achieve. There's pain in there, because it involves ruling some things out. The upside is that you get better at what you do: you get more specialised. **Roger Federer isn't also a professional swimmer –** and in fact Boris Becker reportedly had to choose between playing tennis and playing football, because doing both would have created success in neither. So again, we're here trying to avoid **opportunity cost** – you'll get more done by being choiceful.

Sometimes I hear foundations who don't seem to mind what their process is, and it reminds me of this:

Alice went on: 'Would you tell me, please, which way I ought to go from here?'

`That depends a good deal on where you want to get to,' said the Cat.

`I don't much care where--' said Alice.

`Then it doesn't matter which way you go,' said the Cat.

There's a Dilbert cartoon of a guy answering the phone saying yes certainly'. It's labelled 'Company With No Strategy'. The next cartoon is labelled 'Company With Strategy' and shows him saying 'no we don't do that'.

Again, people in philanthropy say all the time that funders and charities need clear and choiceful strategies, and yet we don't do it. A good test is **put your mission into one tweet**.

Moving now along the process flow....
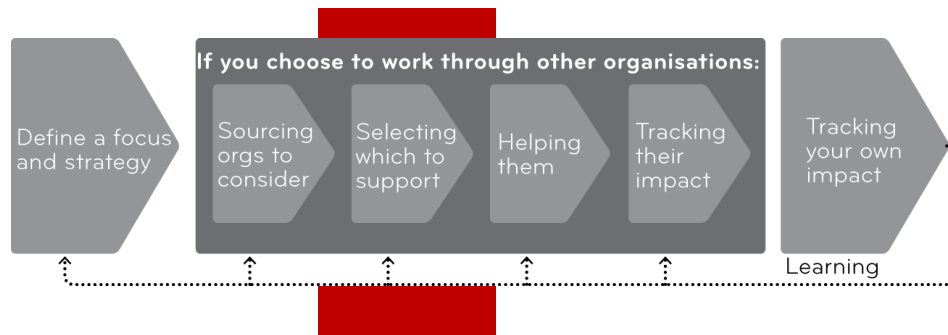
... and deciding which programmes to run, or to fund, here's a useful framework:

$$\textbf{Impact} = \textbf{idea x implementation}$$

That is, your impact depends on two things: your idea and your implementation. They both need to be good.
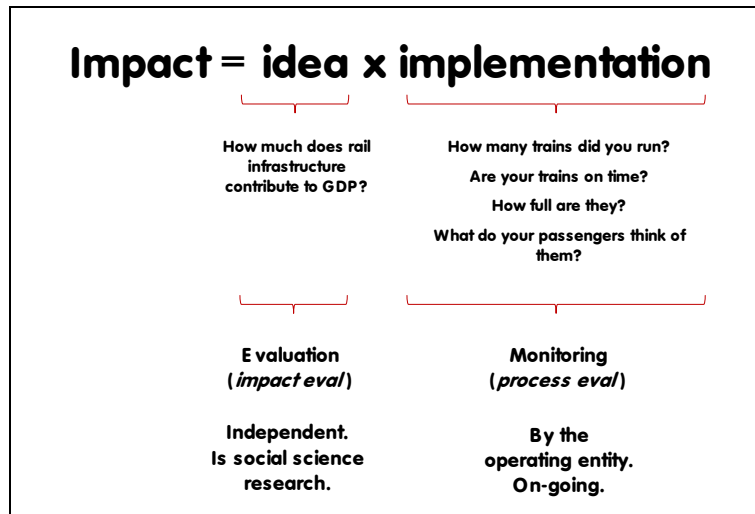
Your idea is your intervention. For example:

- Trains: shall we build more rail infrastructure

- Breakfast clubs: for school children to get a decent breakfast so they concentrate better in class

- Under-arm serves: it doesn't matter how good your volley is when you have to receive a serve.

We can see now the point that both the idea and implementation need to be good. You could be great at implementating trepanning (drilling holes in people's skulls to let the evil spirits out) but since we know that trepanning is a bad idea, this won't help the patient to feel better.
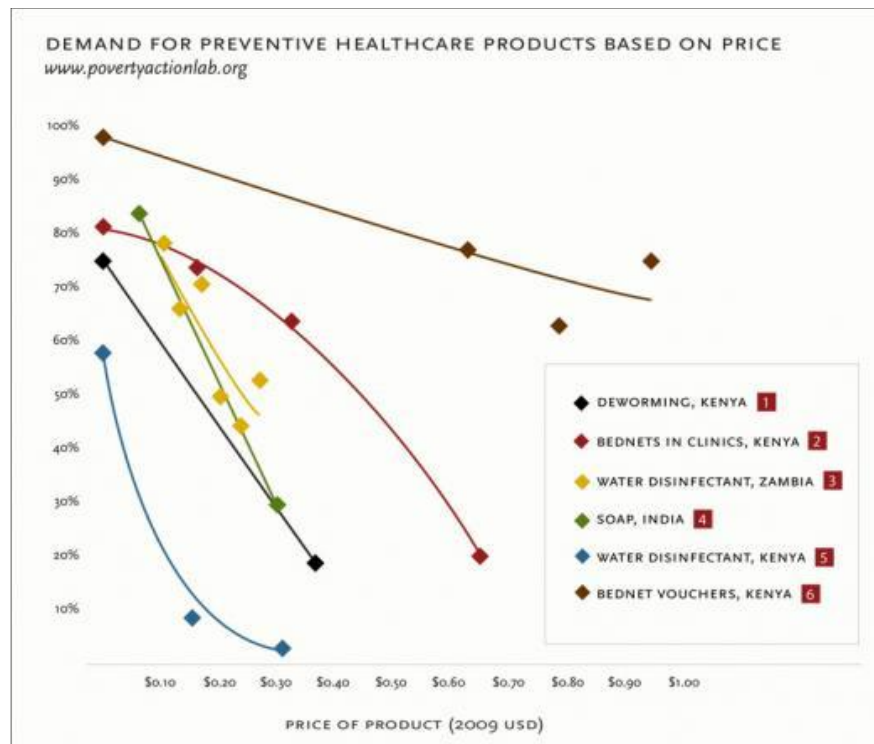
Measuring the effectiveness of implementation is pretty easy. Take the example of trains. Here's we're just asking how many trains you ran, whether they were on time, how many passengers were carried, what the passengers thought of them, and so on. This, the implementing agency (here, the train company; and by analogy the charity) can measure themselves. This is **monitoring**. Implementation is improved by process evaluation, e.g., figuring out how to improve the trains' punctuality, cleanliness etc.

Measuring the effectiveness of the idea is much harder. On the example of trains, we'd be asking whether rail infrastructure affects GDP growth and the scale of that effect. It involves isolating variables to search for causes. **This is impact evaluation**, and it's social science research, e.g., long-term studies, meta-studies, randomised control trials (RCTs). Identifying causes is hard, and hence **the operating charity shouldn't do it themselves**. It needs to be done by people specialised in social science research, and who are independent (remember how charities' work on understanding impact being co-mingled with their marketing.)

Impact = idea x implementation

How much does rail infrastructure contribute to GDP?

How many trains did you run?

Are your trains on time?

How full are they?

What do your passengers think of them?

Evaluation
(*impact eval*)

Monitoring
(*process eval*)

Independent.
Is social science research.

By the operating entity.
On-going.

Fortunately, **the 'idea' will often have already been studied**, so you don't have to guess. Hence some foundations now do a literature review when assessing a grant opportunity. For example, suppose somebody wants to distribute bednets, and to charge for them. Is that a good idea? Well you can just look that up, because J-PAL's tested it umpteen times. Same for soap, and deworming, and other preventative health products (see below). It turns out that charging for these products is a pretty bad idea because the price deters loads of people: by the time you've gone to all the bother of developing a bednet and distributing it, you might as well dish it out for free because charging will stop it being used.

Does that mean that we should never charge for these things? Well not necessarily. But these studies show a pattern, and in the absence of anything else, we'd expect that pattern to hold. The Dutch have a good phrase which is, roughly, 'comply or explain': you can ignore the pattern if you want, but the burden is on you to explain and defend why you've done that.

DEMAND FOR PREVENTIVE HEALTHCARE PRODUCTS BASED ON PRICE
www.povertyactionlab.org

DEWORMING, KENYA [1]
BEDNETS IN CLINICS, KENYA [2]
WATER DISINFECTANT, ZAMBIA [3]
SOAP, INDIA [4]
WATER DISINFECTANT, KENYA [5]
BEDNET VOUCHERS, KENYA [6]

PRICE OF PRODUCT (2009 USD)

There is quite a lot of this research around:
- on international development, look at material from the [Abdul Latif Jameel Poverty Action Lab](#) at MIT (J-PAL) and [Innovations for Poverty Action](#), which have collectively run some 400 impact evaluations. Also the [International Initiative for Impact Evaluation](#)
- on health interventions, look at [The Cochrane Collaboration](#)
- on healthcare in [disaster and emergency situations](#) (earthquakes, tsunamis etc.), look at Evidence Aid, a specialist part of The Cochrane Collaboration.
- on behaviours, look at behavioural economics e.g., this great paper by [Ideas42](#).
- on first world education, look at this toolkit from the [Education Endowment Foundation](#).

If the idea has been studied, then we don't need to study it again. In healthcare, **"not every hospital needs to be a test site"**. The ideas get tested centrally and then just used. So we don't need every charity to be evaluating its idea (doing impact evaluation) – it's complicated and often unnecessary. In those cases, all you need to find out is whether things are going okay, whether anything's changed in the external world, and whether implementation is going okay. That's it.

If we **don't know whether the idea works**, then find out. Properly and rigorously. A tenet of good clinical trials is "ask an important question and answer it rigorously". If we don't have an important question or the money to answer it properly, don't bother; and if it's been answered already, then don't waste time and money researching it idea again. But if the question persists, then do research it, and publish the answer.
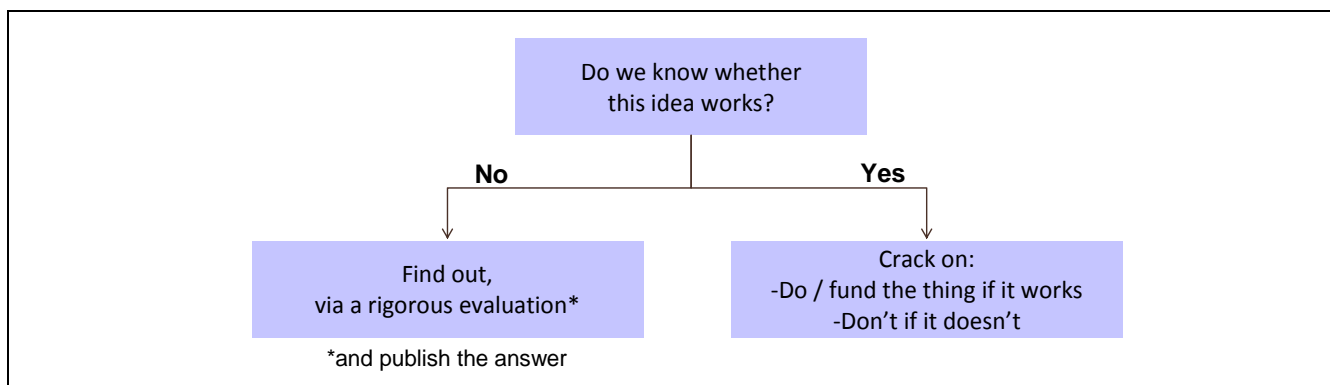
(Giving Evidence will soon publish an article on this topic.) *[Then we split into groups and did some work.]*

Data collection

Let's now think about the quality of data collected. This is important because we're trying to use evidence and evaluation to be brilliant: data is our raw material so it needs to be good. The data is often problematic. **Almost always, the charity is asked to get the data**. Four points on that:

    1. Data collection:

It's totally silly to expect the charities to be able to come up with decent information on their **idea** and to analyse it. There's no incentive for them to do this and funders ask them to **social science research**, the skills for which aren't co-located with good delivery. ([This article](#) discusses the difficulties of good evaluation in relation to one project, the Peterborough Social Impact Bond.)

So **get an independent skilled outsider**, e.g., a relevant academic. The outsider's research will be of interest to: (a) the charity, which should be hungry to learn to improve, and (b) **the wider world**, so publish it.

2. *Data quality*

It's rare for the quality of charities' data to be examined. The Paul Hamlyn Foundation recently published a report which looked at it. It found that charities' data are pretty bad. Only a quarter of reports which PHF received contain 'good' quality data, and some include claims which are totally unsubstantiated:
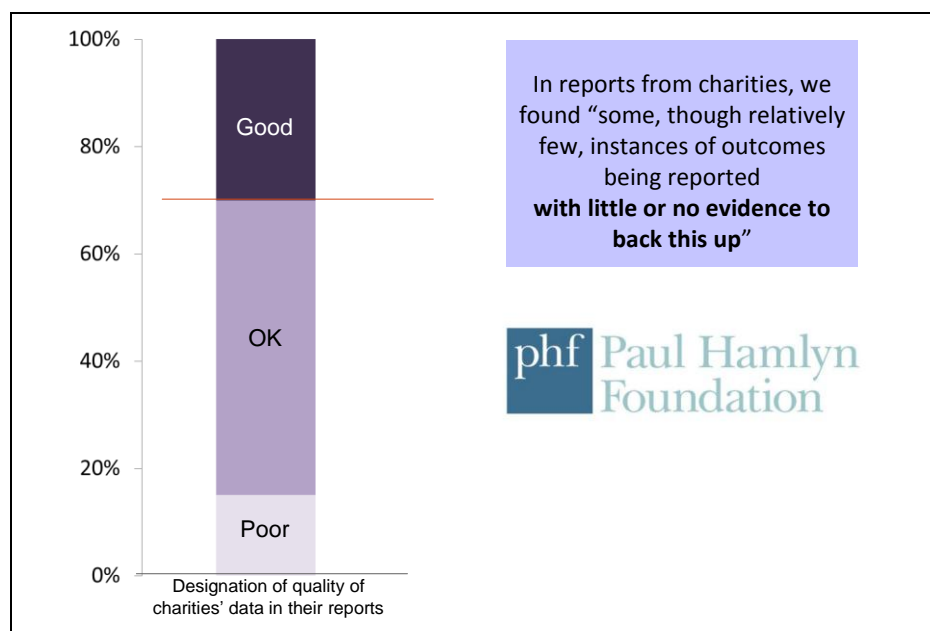
Having been a charity CEO, this doesn't surprise me at all. Producing great quality data is expensive and isn't rewarded at all, and showing evidence of impact also isn't really rewarded (in my experience). Funders rarely ask good questions about the data, so you might as well just make wild optimistic claims. This isn't charities' 'fault': they're just responding rationally to badly-designed incentives.

3. *Tesco principle*

Collecting data which doesn't help the funder or the charity know **what to change in future** is pointless and a waste of time. **Don't ask for stuff just because you can:** funders do this all the time and it's hugely expensive, costing easily £1bn for UK charities alone. [Giving Evidence is working on this.]

Tesco apparently has a principle that HQ isn't allowed to ask a store for any data which isn't useful to the store. This isn't about cost – and certainly not about ethics, which does come into play in charities. Rather it's about data-quality: if the charity has no interest in the data, you'll just get **awful quality data**.

For example, the UK National Health Service recently was concerned about the number of doctors nearing retirement age, so surveyed them all about their age. An astonishing proportion were apparently born on January 1st, 1901! Obviously they weren't, but that was the first item in the drop-down list, and doctors had no reason to give accurate data and just wanted to get the survey over and done with.

4. *Proportionality*

How much work should the monitoring and evaluation be? The aim is to ***find out something we don't know***. If we know about the idea, we don't have to test that.

It's often said that the evaluation of a grant (or, more precisely, of the work enabled by a grant) should be proportionate to the size of the grant. It's also often said that evaluations should be proportionate to the size of the charities. We can see now that both views are wrong. The aim of an evaluation is to generate new and useful knowledge, and hence **the amount worth spending on an evaluation is proportionate to the size of the knowledge gap, and the scale of the programmes which might use that knowledge**.

The right analysis is the cost and benefit of the new knowledge.

To illustrate, suppose a new drug for breast cancer has been developed by a small company. The 'first (wo)man in' trials, involve only few people, for safety reasons. Relative to the cost of giving the drug to those 'first women in', how much should be spent on evaluating what happens? 'A lot', obviously, because the knowledge could be used at significant scale. The size of the company or the cost of administering the drugs are totally irrelevant. Similarly, it's worth spending a lot evaluating a small pilot of a new idea for improving prisoners' literacy. Often, the cost of robustly evaluating a programme will exceed the cost of delivering that programme – and that's fine if and only if the results are useful to a much wider audience.

Consequently, if we don't have a knowledge gap (because the idea is already studied), we don't need new data, so the cost will be very low.

And don't ever collect data you don't need. That sounds perfectly obvious but it goes on all the time!

Conclusion

Bills Gates said that "measurement saves lives". However, for that you need to **collect what you need, use the data you need, and don't collect what you don't need. This is how evaluations can make you brilliant.**

> ✓See what exists already: don't duplicate
>
> ✓Collect what you need
>
> ✓Collect it properly to properly answer the question
>
> ✓Don't collect what you don't need

Giving Evidence is a consultancy and campaign, which supports donors to give based on sound evidence. Founded by Caroline Fiennes, author of acclaimed book *It Ain't What You Give* and award-winning charity CEO, we work with foundations, individuals, families and companies on using evidence to improve performance.

Giving Evidence