

# ‘Real World Evaluations’ – randomising on a national scale

by **Clara Marquardt**, posted on Oct 8th 2013

*Giving What We Can promotes effective altruism - giving informed by rigorous evidence of programmes' and projects' effectiveness. An increasing demand for evidence necessitates a corresponding growth in the supply of such evidence by NGOs & social enterprises. Clara Marquardt from the Cambridge GWWC chapter interned over the summer with Development Media International (DMI) (<http://developmentmedia.net>), an organization aiming to conduct "the most robust evaluation that has ever been conducted of a mass media intervention in a developing country". Clara took the opportunity to discuss the movement towards impact measurement and the challenges associated with real world randomized controlled trials with Will Snell, DMI's Director of Public Engagement & Development.*



**Clara: Could you give us a brief overview of DMI's objectives and approach?**

Will: I would describe DMI as an organisation that uses radio and television to improve health outcomes, especially maternal and child health outcomes, in developing countries by changing behaviours. We are bringing together two different disciplines. One is the creative and logistical expertise in managing large-scale multi-issue media campaigns focused on for example the prevention and treatment of diarrhoea or safe child birth; the second is the scientific rigour of measuring the health impact of those campaigns.

We have developed a statistical model with the London School of Hygiene and Tropical Medicine (LSHTM) (<http://www.lshtm.ac.uk>) that allows us to predict and then measure not just behaviour change outcomes of a campaign but also health impacts in terms of numbers of lives saved. Historically there has never been any such attempt to measure or model the impact of media campaigns on mortality. This is because the impact on all-cause mortality is very hard to identify statistically given the many possible causes of deaths and the limited impact of any campaign.

The model predicts that DMI's campaigns will reduce under-five mortality by 16-23% within each African country at a cost that is as low as that of any other available intervention...between 2 and 10 USD per DALY.

**Clara: You have already mentioned a defining characteristic of DMI - a strong focus on 'science' and impact measurement. How did this come about?**

Will: This is a story about Roy Heads, DMI's founder and chief executive who had previously set up and run the BBC World Service Trust's health division and in doing so had worked on a large number of extremely successful mass media health campaigns. He set up DMI with the objective of injecting a more scientific approach because he felt that the impact of these campaigns while significant was not being measured sufficiently robustly.

Amongst other things this was leading to a lack of prioritization of such campaigns by funders and governments...the aim was to prove their impact and thereby persuade donors to make this a public health priority alongside complementary supply side initiatives. The scientific evidence base could also be used to maximize a campaign's mortality impact by informing the choice of health issues to campaign on.

**Clara: Is DMI then starting to see this strategy pay off? Are donors and funders responsive to the provision of rigorous evidence? Which type of funder is proving the most receptive?**

Will: DMI is starting to see the benefits and an increasing interest among funders. At the moment we have a track record, an evidence base based on behaviour change outcomes. The first time that we are directly measuring mortality outcomes is in the current Randomised Controlled Trial. While that is on going (endline results will be published in 2015) we don't have hard evidence on health outcomes other than a couple of campaigns where have been able to measure trachoma prevalence - in Ethiopia for instance. A lot of funders are responding to what we are doing. By definition the most science driven funders are waiting for the results of the Randomised Controlled Trial - this is gold standard evidence for them.

I would say that the major push for impact measurement is coming from certain bilateral donors who are extremely evidence and research focused, and also from some more public health focused organisations. Both groups of donors are driven partly by value for money considerations... I agree that the development of new financing mechanisms such as Development Impact Bonds (DIB) and the movement towards impact measurement are mutually reinforcing. The development of a DIB market is dependent on a certain amount of accurate data and evidence being available but once DIBs become more widespread this will

motivate NGOs and other organisations to become more robust in terms of impact measurement in order to secure continued funding.

**Clara: You have already mentioned DMI's on-going Randomized Controlled Trial (RCT) - an evaluation quite unique in its scope and scale. How would you describe the project?**

Will: We were in discussions with the Wellcome Trust when developing the DMI-LHSTM model to predict how many lives we save and they wanted us to provide some real world evidence that this could work. They funded us to set up a RCT to provide gold standard evidence that observed mortality reductions are attributable to our campaigns rather than other factors.

This is an evaluation design that has rarely if ever been used in the field of media - it is very difficult to randomize media interventions. We looked at nearly every country in Africa and Burkina Faso was the only country where this design could work because radio is very fragmented and localized. While there is a national network very few people listen to it, partly because it is in French, partly because it is under resourced. Everyone listens to local FM community radio stations and TV is more or less non-existent in rural areas. Given the presence of 'media dark areas' we were able to select 14 zones across the country each measuring roughly 50 km radius around a small town and randomly selected 7 to be intervention zones and 7 to be control zones.



Map of Burkina Faso showing intervention areas in blue and control areas in red

In the intervention zones we are broadcasting a 2.5 years child health campaign - a mixture of radio spots (one minute adverts broadcast 10 times a day) and dramas (interactive 5-10 minute health dramas broadcast every evening) (see [here](http://developmentmedia.net/video-and-audio-library) (<http://developmentmedia.net/video-and-audio-library>) for some examples of spots & dramas).

**Clara:** From a technical perspective, you have already explained how you managed to implement treatment randomisation on a national scale. Are there any other areas where the Burkina RCT did diverge from the idealized

**'textbook' RCT methodology? In other words, what are the challenges of actually realizing this rigorous evaluation design?**

Will: Sure, the most obvious way in which we have diverged from an idealized design is in the number of clusters. Having 7 intervention and control clusters reduces our statistical power to attribute impact. There are obvious practical reasons why we couldn't go beyond that: There are a limited number of radio stations in Burkina Faso with a sufficient listenership. Also, this is a 12 million USD trial - would we have doubled in size we would have more or less doubled in costs. That notwithstanding we have a 80% statistical power (a one in five chance that we will have a significant impact on mortality but won't be able to prove it statistically).

The other area in which we have faced scientific challenges has been around controlling for other health interventions. There are a number of nationwide interventions including vaccination campaigns that should theoretically affect control and intervention clusters equally. But there have been some health campaigns where there has been a possibility of them broadcasting or operating in some areas but not in others. We have been working closely with the Ministry of Health to try and avoid that happening during the broadcasting period of our trial. That has been possible in most cases so we think that contamination from other projects is minimal... but will never be completely eliminated.

**Clara: You have referred to two potential constraints on the implementation of RCTs: Their high cost and their long duration which makes it harder to minimize potential contamination. Based on such considerations what is your take on the World Bank's assessment that "experimental designs...can only be applied in a very small proportion of impact evaluations in the real world...RCTs can probably only be applied in less than 5% of impact evaluations" [1]? Seeing as DMI has already shown that randomization, one**

**of the key challenges, is realisable on a national scale do you see the Burkina RCT as a confirmation that RCTs can be become more widespread?**

Will: We knew this from the start - we would not be expecting to run RCTs on a regular basis. We will look at opportunities for future RCTs because it is clearly a level of rigor that other designs will not achieve ... but it remains difficult to run RCTs around media in most contexts.

For campaigns in other countries we will probably be looking at quasi-experimental evaluation designs such as non-randomized controls and time series analysis which permits a much greater degree of flexibility in terms of campaign design.

**Clara: Would you nevertheless agree that there are ways of reducing the cost and burden of conducting RCTs? Would you in this sense see your strategic collaboration with LSHTM as a model for future RCTs? How about the call for 'piggybacking' on existing data gathering efforts in place of conducting independent and costly baseline and endline surveys?**

Will: I do think the bringing together of DMIs and LSHTM's technical expertise is a viable model. Our future research will no doubt be based on similar partnerships. Any partnership brings its own complexities but I think that the benefits to this kind of partnership outweigh these. From the NGO, implementer perspective it makes sense to partner with an academic body with the independence to provide credible results and the necessary technical Measurement and Evaluation expertise. From the academics' perspective, the NGO will hopefully have the expertise in managing and implementing the intervention in question.

I think that there are increasing possibilities of using secondary data sources, especially when a project is in some way interfaced with the health system and

outcomes are measured in terms of increased health system usage (which is measured by Ministries of Health) - appointments for children with malaria or sales of Oral Rehydration Solution (ORS) sachets for instance. It has been difficult for us in Burkina Faso partly because there are few reliable data sets and partly because many of the behaviours we are promoting are household behaviours such as breastfeeding that the government is not recording. RCTs demand a level of rigour at every stage of the process that limits the number of appropriate existing data sets.

**Clara: Are there any challenges associated with DMI's movement between RCT and quasi-experimental evidence? In other words, is there a risk that once DMI is committed to the RCT standard of evidence funders will discount other evidence?**

Will: There certainly is an issue around potential funders waiting for the RCT results despite the fact that our evidence base is already stronger than that associated with most other (mass) media campaigns. In terms of moving from RCT to non-RCT evidence in the future we can break it down into two parts: One is the attribution of impact and the second is the conversion of behaviour change outcomes into lives saved impacts.

The RCT's aim is to as far as possible show that firstly, there is a direct link between the radio campaign and lives saved and secondly, that our model accurately predicts lives saved based on the intermediary behavioural change variable. Assuming that this is the case we will then be able, with reasonable rigor, to say that in future campaigns we don't need to measure mortality which is much more costly for a numbers of reasons including the larger sample sizes required. Instead it will be sufficient to measure behaviour changes and use the [Lives Saved Tool \(LiST\)](http://www.jhsph.edu/departments/international-health/centers-and-institutes/institute-for-international-programs/list/) (<http://www.jhsph.edu/departments/international-health/centers-and-institutes/institute-for-international-programs/list/>) (an independent tool funded by the



Bill & Melinda Gates foundation that predicts how many lives could be saved in different countries if the coverage of key interventions such as breastfeeding and insecticide treated bed nets was increased from their current level) to translate behaviour changes into lives saved.

The remaining challenge then is to attribute our behaviour change impact to the campaigns. This will require us to use the most sophisticated quasi-experimental designs that strike the right balance between robustness and technical viability - we are working on developing such kinds of evaluations. One example would be dose response analysis: When we are conducting Knowledge, Attitude and Practice surveys (the standard term for surveys that assess people's belief and behaviour changes) we also ask people whether and how much they have been exposed to the campaign. We can then break down the behavioural data by campaign exposure. If people with low campaign exposure changed their behaviour less than people with high exposure that provides us with the additional evidence that the campaign itself was having the impact.

**Clara: There has always been a concern over the external validity and generalizability of RCT results. Specifically, a concern over the lack of understanding of the process of change behind observed results and the contextual, mediating factors that enabled the results to come about. In DMI's case these could include supply side health interventions or specific features of the radio stations' broadcasting formats. What is your understanding of these factors in relation to the Burkina RCT? To what extent can you move beyond the statement that "a very specific mass media campaign using context specific stories reduced mortality in the specific socio-political environment of Burkina Faso, Africa by x%"?**

Will: The longer and more complex your theory of change the harder it is to generalize the results of an RCT which reveals significant impacts of an

intervention. Our theory of change is relatively straightforward with effectively only one intermediary variable and outcome - behaviour.

In order to better understand the link between media messages and behavioural changes we are conducting qualitative research to compliment the quantitative research. We do this partly to inform our message production in terms of understanding barriers to behaviour change and to pre-test spots for accessibility and impact but also to understand responses of our audiences using feedback research with individuals and focus groups. That gives us some idea - anecdotal - to what extent people are changing their behaviours and whether if they are changing them that results from direct exposure to the campaign or from having heard the messages being discussed in social forums, from neighbours and friends.

**Clara: Everyone within DMI is very aware of the limitations of the trial's results but how do you communicate this to the public? How will DMI communicate the results so as to maximize their (policy) impact while acknowledging their limitations and contextual background?**

Will: If the results are positive of which we are confident - but then you can never guarantee this - the challenge will be to be optimistic but realistic about their generalizability and contexts. We are optimistic about this. While Burkina Faso is an easy country to work in in many ways it is also a very challenging environment in terms of healthcare provision and media penetration. Programs that work in context of Burkina Faso's very limited health care service are likely to work in other countries.

The opposite danger is that the results foster a sense that any media campaign can have a similar impact. We will be doing our utmost to emphasize that much of the success of the trial was due to the way in which it was implemented. Promoting not just the use of media to change health behaviours but a specific approach as to

how to do that. We have developed an approach which we call Saturation+ that comprises three aspects: Saturation, reaching the largest possible proportion of our target audience; science, using the models we discussed to maximize health impacts and to measure health impacts robustly; and stories, investing in both the qualitative research to identify behavioural barriers and a creative process linking local scriptwriters with international experts and expertise.

**Clara: A recent publication on Development Impact Bond has called for them to be "open by design", to use a "share standard of reporting and conducting trials", and to "create organizations and funds to...systematize the evidence"[2]. To what extent are these observations relevant to RCTs? In what ways will DMI contribute to the building of a body of evidence on how to conduct real world RCTs? Is there a potential conflict between DMI's social motivation and purpose which is served by growth of evidence-based approaches and the need to protect DMI's expertise and distinctive features?**

Will: The DMI-LHSTM model has an intellectual property aspect to it, but all the model does is to model health impacts and thereby maximize campaign impact via message selection. The most effective way to maximize health impacts lies in the approach to running campaigns - our communication strategy will therefore focus heavily on our Saturation+ approach that we will be sharing widely.

Our focus is on sharing the results and the information that helps people run effective media health campaigns rather than on sharing insights on how to run effective RCTs. This goes back to our view that there most likely won't be too many future RCTs in this area. Where forums such as the International Initiative for Impact Evaluation (3iE) or the Poverty Action Lab (J-Pal) are however interested in aspects of our implementation process we will share this information.

**[1] "Conducting Quality Impact Evaluations under Budget, Time and Data Constraints", The World Bank Independent Evaluation Group, 2006**

**[2] "Development Impact Bond Working Group Report", Centre for Global Development and Social Finance, 2013**

*This is a slightly compressed and edited version of the original interview. Thanks to Will Snell for participating in this interview!*

**Clara Marquardt**

◀ Previous Post [\(/blog/2013-10-01/interning-at-giving-what-we-can-the-peoples-perspective\)](/blog/2013-10-01/interning-at-giving-what-we-can-the-peoples-perspective)  
Random Post [\(/blog/2012-10-27/estimating-the-effectiveness-of-dcp2\)](/blog/2012-10-27/estimating-the-effectiveness-of-dcp2)  
Next Post ▶ [\(/blog/2013-10-16/g8-keeping-promises-a-look-at-the-lough-erne-accountability-re\)](/blog/2013-10-16/g8-keeping-promises-a-look-at-the-lough-erne-accountability-re)

**2 Comments**   **Giving What We Can**

**1 Login** ▼

♥ Recommend   [↗ Share](#)

**Sort by Best** ▼



Join the discussion...

LOG IN WITH

OR SIGN UP WITH DISQUS

Name



**Anonymous** • 4 years ago

I work in the field in West Africa and ask - Do we really want to abandon potential opportunities to relieve suffering in impoverished populations just because methodological experts don't yet have the perfect tool for the conditions we are dealing with?

^ | v • Reply • Share ›



**Anonymous** • 4 years ago

As Will rightly points out, having only 7 intervention and 7 control clusters significantly reduces the statistical power of the study's impact assessment tests. Notwithstanding his acknowledgement, this makes me doubt the extent to which this is "the most robust evaluation that has ever been conducted of a mass media intervention in a developing country". There is a reason why RCTs are rarely if ever used in the field of media - it is difficult to identify a sufficient number of independent clusters to make the estimation results sufficiently credible. This is exactly why wise evaluators decide not to go for this approach. Yes - RCT evaluations certainly bring in money because many donors are not aware of power calculations and only know the 'RCT = golden standard' rhetoric. But do we really want to give up on true learning opportunities just for the sake of pleasing uninformed donors? I am getting increasingly worried about the fact that NGOs seem to be moving from the 'RCT is equal to horror' to the 'RCT is synonym for rigor' discourse, ignoring the fact that RCTs - even if indeed being best practice in many circumstances - might not always be the most suitable evaluation design to follow given the constraints of project design.

^ | v • Reply • Share ›

#### ALSO ON GIVING WHAT WE CAN

### What's so special about GiveDirectly's basic income pilot?

2 comments • a year ago

**Brandy Moore** — Many people given a basic income do start businesses. The idea is that a basic income will ...

### It's Good to Talk - the Giving What We Can Blog

1 comment • a year ago

**Adam Millest** — Great article!

### Is deworming really effective? A response to recent deworming studies

4 comments • 2 years ago

**Frikster42** — Maybe they're understaffed. It's not out of the question. A lot of non-profits often have one of those two

### The Economic Benefits of Malaria Eradication - the Giving What We Can ...

3 comments • 2 years ago

**paulo rufalco moutinho** — Hello, I'm entomologist and work with Amazon malaria. I'm writing a paper and I like to