PRO BONO ECONOMICS

MO FARAH FOUNDATION

Evaluation advice for the Mo Farah Foundation

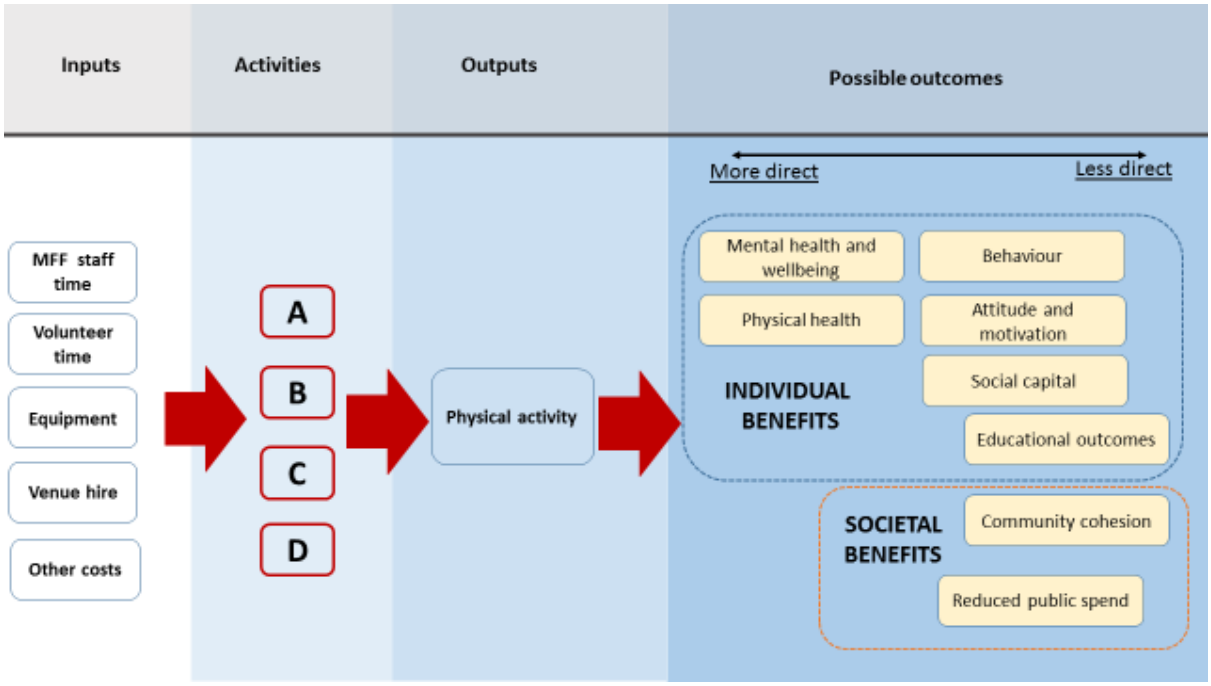
October 2014

Evaluation advice for the Mo Farah Foundation

Executive summary

The Mo Farah Foundation is seeking advice on how to evaluate a programme of interventions aimed at increasing physical activity among certain target groups. The primary objective of the programme is to improve the physical and mental health of participants. There may also be less direct effects on pro social behaviour, attitudes and motivation, levels of social capital and, ultimately, education outcomes. There may also be benefits for society at large, through reductions in public spending (on health and, less directly, welfare payments) and stronger community cohesion. These impacts are set out in the logic model in Figure 1.

Figure 1: Possible logic model, linking intervention to final outcomes



Ideally a full impact evaluation would be able to measure the *additional* impact of the programme on the full range of outcomes, and assess whether there is a causal link from the programme's activities. Where possible, monetary values would be assigned to the benefits, which could then be compared to the full costs of the programme.

As is discussed in more detail below, in this case such an evaluation would be expensive, is unlikely to fully succeed and is probably disproportionate for the expected size of the programme. We recommend that a more sensible approach is to focus on measuring the key change the programme is seeking to achieve – enhanced physical activity – and then to apply estimates of the value of these change from other evaluations. This would require the programme to capture the following information:

- Detailed information on who takes part in which activity
- The unit costs of providing each activity
- The regularity, duration and intensity of the activities undertaken

- The persistence of the effects on behaviour *during and after* the programme has taken place.
- Socio-economic data on the participants to enable a comparison with those who are not taking part in the programme

The inputs and activities of the programme

Specifying the Intervention

MFF seeks to fund a range of different activities within a given programme. There is an important decision to be made about whether to evaluate the programme *as a whole* or the different components, or 'treatments' within it. The latter approach requires more information and will be more costly and time consuming. However it is potentially much more useful as it will enable judgements to be made about which elements of the programme are most effective and what provides the greatest impact relative to its cost. This in turn should help to improve MFF to improve the impact of their programme over time, and to advise other organisations doing similar work.

In order to compare across treatments, MFF will need to: (1) specify what the interventions are; (2) collect information on who is subject to each, and how often, and; (3) calculate unit costs for each.

If it is *not* possible to separate which individuals took part in which activities, and which did not, then it will not be possible to say which activity is leading to any change that is observed.

Costs

It is important to make an effort to estimate the costs of providing the different activities, so that a judgement can be made about the cost-effectiveness of each one. Over time this will enable more resources to be directed towards the activity that offer the best return. Some costs can be allocated to specific interventions and the primary cost is likely to be the time of those who organise and run the activities. Effort should therefore be taken to record the details of the time spent – by staff and volunteers – on each activity where possible. There will be some fixed costs that are shared across the whole programme.

Programme outputs

While the programme ultimately aims to improve a range of individual and social outcomes, the immediate output is an increase in physical activity among programme participants.

To attempt any evaluation, the programme must collect accurate measures of who is taking part, how often and for how long. As discussed below, previous studies have found that the benefits from physical activity depend on the *frequency*, *duration* and *intensity* of the activity. So it will be important to find ways to capture these aspects.

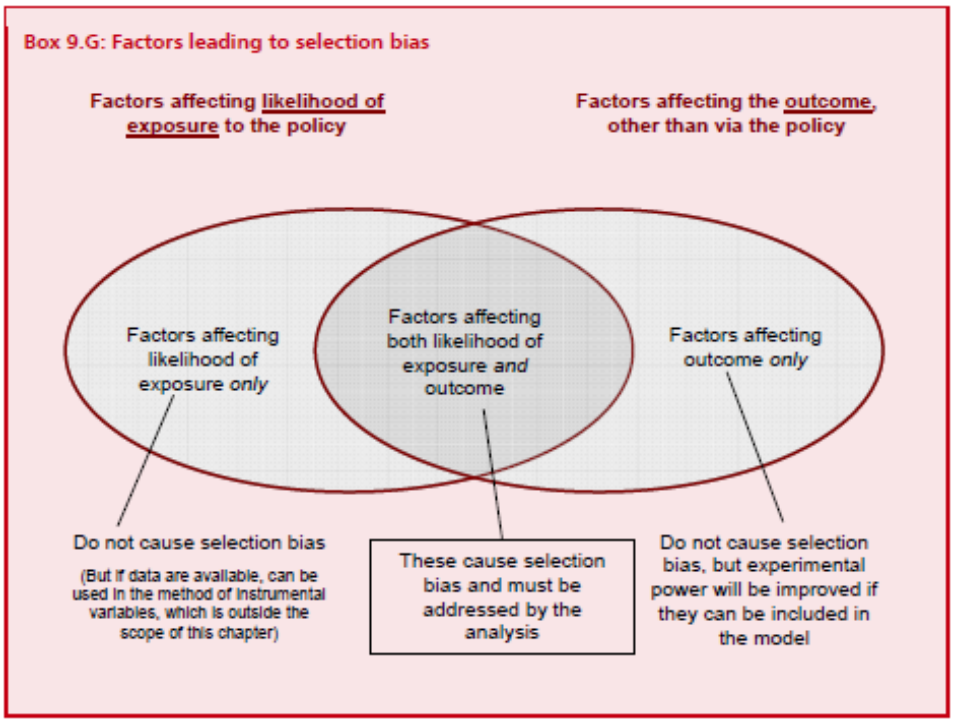
Assessing the additionality of the programme

To understand the additional impact of any project or programme, it is important to estimate what would have happened without the intervention. Constructing this 'counterfactual' is a key part of

successful impact evaluation.[1] In this case, we want to know the level of physical activity that participants would have done in the absence of the programme.

For the comparison to be valid, we need to be able to compare like-with-like. That is, we need a valid control group, who are similar to those taking part in the programme. If there are factors which affect both the likelihood of an individual taking part in the programme, and the outcome, this will bias the results. In simple terms, it will not be possible to tell whether any differences we observe are due to the intervention, or to the characteristics of the person who is undergoing the intervention (Figure 2). Any evaluation results will be severely weakened.

Figure 2. Factors leading to selection bias[2]



This is a considerable risk for this programme. MFF's programmes are intentionally targeted at certain groups, and where the propensity to exercise varies across these groups, there is be scope for bias. In addition, those who *volunteer* to take part in the programme are probably more likely to be motivated to do physical activity regardless of whether they are in the programme. In addition, if

There are a few ways that an evaluation might deal with this and more accurately compare like-with-like.

The 'gold standard' would be to randomise the intervention. That is, from within the population of possible participants, to randomly choose who takes part and who does not, and then to compare between the two groups. A variant on this would be to roll out the programme in phases and to randomly assign individuals to different phases. The outcomes of participants in the first phase could then be compared to the outcomes of the future participants of further waves. The strength of these studies will depend on whether the two groups are representative of the population of interest,

[1] *The Magenta Book*, paragraph 9.12
[2] Taken from *The Magenta Book*

whether the sample size is sufficiently large and whether randomisation has been conducted appropriately.[3]

Another approach is 'matching', where a technique such as *propensity score matching* is used to select subsets of participants and non-participants who have similar characteristics. Simple comparisons can then be made between the two groups.

Another strategy is to use statistical methods to compare the outcomes of participants with those of a dissimilar group, and to control for the different characteristics using regression analysis. Any remaining policy effect can be interpreted as a causal effect of the policy.[4] It should be possible to make use of existing surveys of physical activity, to compare activity levels among individuals of similar ages (for example, the *Taking Part* survey).[5] This will require MFF to record the relevant socio-economic variables of participants, so that a future evaluation can control for these characteristics.

A counterfactual will be necessary if a future evaluation is to robustly assess a causal link between the programme and improvements in physical activity.

We recommend that the evaluation makes use of the Taking Part survey to estimate additionality using regression methods. The Taking Part survey also collects a range of background data on individuals and a 2011 report analysed the characteristics that help to predict participation rates among children aged 5-10 and 11-15.[6] The sample size was 2,622. This analysis can be used to understand which background characteristics MFF would need to collect in order to be able to make valid comparisons with their target group. The research found the following characteristics to be important factors:

This means the programme will need to collect information on the participants that helps to predict how much physical activity they will undertake. Based on analysis of the Taking Part survey key participant information to collect is:

- Child age
- Gender
- Ethnicity (the analysis only distinguished between white and BME)
- Highest qualification level of the Household Reference Person[7]
- Employment status of HRP – full time, part time, none
- Household income
- Access to a car

[3] Campbell and Harper (2012) *Quality in policy impact evaluation: understanding the effects of policy from other influences (supplementary Magenta Book guidance).* HM Treasury
[4] For a more detailed discussion, see chapter 9 of the Magenta Book.
[5] More detail on the Taking Part survey is available here
[6] Jones, Millward and Buraimo (2011) *Child Participation in Culture and Sport. Analysis of the 2008/09 Taking Part survey*. This report for DCMS can be found here
[7] Approximates to the 'head of household'. Qualification levels were classified as: [1. Higher Education; Other Higher Education below degree level; A levels, vocational level 3 & equivalents; Trade Apprenticeships; GCSE A*-C (5 or more), vocational level 2 & equiv; GCSE (less than 5 A*-C), other qualifications at level 1 and below.

The actual questions asked can be found here. The importance of socio-economic characteristics in explaining participation rates mean that it is important to capture *some* information about this. If it is considered too problematic to investigate the household income of participants, we would recommend asking about highest qualification level – this should provide a good proxy.

Programme outcomes

The MFF programme/s might have a range of benefits for participants and the wider community. Some flow fairly directly from increases in physical activity – for example improvements in physical health. Others are less direct and thus will be more difficult to link to MFF activity. We understand that key outcomes for MFF are improved health and wellbeing among participants.

Health benefits

Improvements in physical health are a valuable outcome in their own right and could provide benefits to the exchequer, via reduced health spending. Some direct health measures could be captured, such as participants' weight or fitness levels. These may be useful measures of programme impact - however at present, these cannot be assessed in monetary terms.

Many of the potential health benefits from exercise occur far in the future, well beyond the likely time horizon for the intervention, for example reduced likelihood of heart disease or stroke. Measuring the actual impacts on these outcomes are likely to be beyond the time horizon of any evaluation.

A common method for placing monetary values on the benefits of particular intervention is to apply estimates of impact from similar programmes. This is sometimes referred to as a 'benefits transfer' method. Applying this method can reduce the evaluation burden considerably and we would recommend it in this case.

The best source for MFF's purposes is likely to be the 2010 report for the Culture and Sport Evidence (CASE) programme.[8] This estimates the health benefits of different types of exercise and at different ages. Tables 14-17 estimate the direct health cost savings associated with taking part in athletics at £1,030 - £2,240 for 11-15 year olds. A further estimate is made of 'total economic value', which includes health cost savings and improved quality of life. This is in the range £6,720 - £14,720.

It is important to understand the reason for the large range in the estimates. The upper bound estimates the impact if the exercise is undertaken 3 times a week at 'moderate' or 'vigorous' intensity. The lower bound estimates the impact if the sport is undertaken at the *actual frequency and intensity* reported in the Taking Part survey. This survey found that the actual activity undertaken in athletics is commonly not at a vigorous or moderate level.

One downside of the benefits transfer approach is that it does not allow you to develop your own bespoke measures. For example, in this case, the CASE report only estimates the impact of exercising at fairly broad brush measures of frequency (once a year or less; several times a year; at least once a month; at least once a week). So there would be no opportunity to test the differential effect of more frequent participation, the intensity of participation, or different types of activity.

[8] Understanding the value of culture and sport (2010). Research report conducted by Matrix for CASE.

We recommend a benefits transfer approach. If it is to be successfully used in this case, the evaluation should seek to:

- Measure the frequency of exercise by programme participants. This has two parts: first, recording the number of sessions that people take part in; second, estimating what is additional to what would have happened anyway (the 'counterfactual').
- Measure the actual intensity of exercise by participants. In order to robustly apply the upper bound estimates, MFF would need to demonstrate that activity was actually at high levels of intensity.

Mental Health and Subjective Well-Being (SWB)

There is growing interest in using subjective well-being (SWB) levels as a direct measure of a programme's success. SWB is recorded through individual self-reports on how happy they feel or how satisfied they are with their life (interestingly these two measures of *affective* and *evaluative* wellbeing are distinct and not fully correlated).

Where a significant difference in SWB can be demonstrated, a monetary value can be assessed by estimating how much additional income would be required to yield an effect of the same magnitude. There are a few challenges to this approach:

- First, monetising the impact is subject to a contested literature on the relationship between income and SWB. Because some studies find a weak link between increases in income and SWB, this means that the equivalent increase in income can be quite large.
- Second, a major challenge with subjective well-being data is that the plausible size of the intervention on SWB is small relative to the natural variation. That is, there is a high ratio of 'noise' to 'signal'. Researchers have typically only been able to explain a small proportion (between 3% and 15%) of the observed variation, even after taking into account a large number of explanatory variables.[9]
- Socio-economic factors mean that average SWB levels vary significantly between different social groups, so considerable effort would be required to construct a robust counterfactual (discussed further below).

The upshot of this is that to have a chance of detecting a statistically significant effect, the evaluation would require a large sample size, and a strong counterfactual (Figure 3).

[9] OECD (2013) *OECD Guidelines on Measuring Subjective Well-being*. See p216

Figure 3 The Importance of research design[10]

	Weak design Poor counterfactual or none at all	Strong design Realistic counterfactual estimate
Low power Small number of observations and / or policy effect small relative to noise	Unlikely to detect difference between groups or over time. And even if we do, we have no confidence in attributing it to the policy.	Unlikely to detect difference between groups. But if we do, then we have confidence in attributing it to the policy.
High power Large number of observations and / or policy effect large relative to noise	Very likely to find a significant difference between groups but this does **not** mean it can be attributed to the policy.	Very likely to find a significant difference if there is a real policy effect. We have confidence in attributing this difference to the policy.

Again, a benefits transfer approach could be adopted, relying on estimates provided in the CASE report. This estimates the SWB impact of frequent engagement in sport as being equivalent to an increase in income of £11,000. It is also worth bearing in mind that the impact of income on SWB is still contested, so the £11k figure could be revised down (or up).

I would recommend that you do not pursue an evaluation based on SWB data, simply because it is unlikely that you will be able to uncover statistically significant results.

Wider benefits

MFF also aim to achieve wider community and social benefits. These are likely to be hard to capture and monetise. If these are an important objective of the project, the first step is to be clear and precise about the nature of the impacts and how the intervention might plausibly affect them. Then consideration could be given to an evaluation strategy. We have not considered in detail at this stage.

Social capital is one possible dimension that could be explored in an evaluation. There are several alternative measures of social capital. These are described, alongside a range of possible social capital questions for use in surveys, in ONS' Social Capital Question Bank. [11] The main measures are:

- **levels of trust** - for example, whether individuals trust their neighbours and whether they consider their neighbourhood a place where people help each other
- **membership**- for example, to how many clubs, societies or social groups individuals belong
- **networks and how much social contact individuals have in their lives** - for example, how often individuals see family and friends

Conclusion

Although it would be possible to design an evaluation to assess the direct health and wellbeing impact on programme participants, this would be complex, expensive and is not certain to be successful. A more appropriate strategy should focus on collecting sufficient data that enables MFF to do two things: first, to understand the cost and impact on physical activity of different

[10] Taken from HMT (2011) *The Magenta Book. Guidance for Evaluation*. Table 9.B.
[11] **Social Capital Question Bank June 2002 - ONS**

interventions within the overall programme, and; second, to apply estimates of the value of enhanced physical activity from similar programmes elsewhere. This will require the following information to be collected:

- Distinctions between the different interventions within the programme
- Expenditure and resources, including staff time, allocated to each intervention
- Which participants are subject to which intervention
- The duration, intensity and frequency of physical activity for each participant
- The persistence of the behaviour change
- The socio-economic characteristics of programme participants

Tom Aldred, HM Treasury / Pro Bono Economics
September 2014