PATHFINDER

Guidance on Outcomes Focused Management

Building Block 4 : Assessing Impact

Version 2.1, July 2003

This document provides guidance for agencies integrating outcome information into their decision-making processes. The document was produced by New Zealand's Pathfinder Project. More Pathfinder guidance documents are available on http://io.ssc.govt.nz/pathfinder.

We hope other outcome-based initiatives continue to develop the material presented in this suite.

Building Block 4 : Assessing Impact

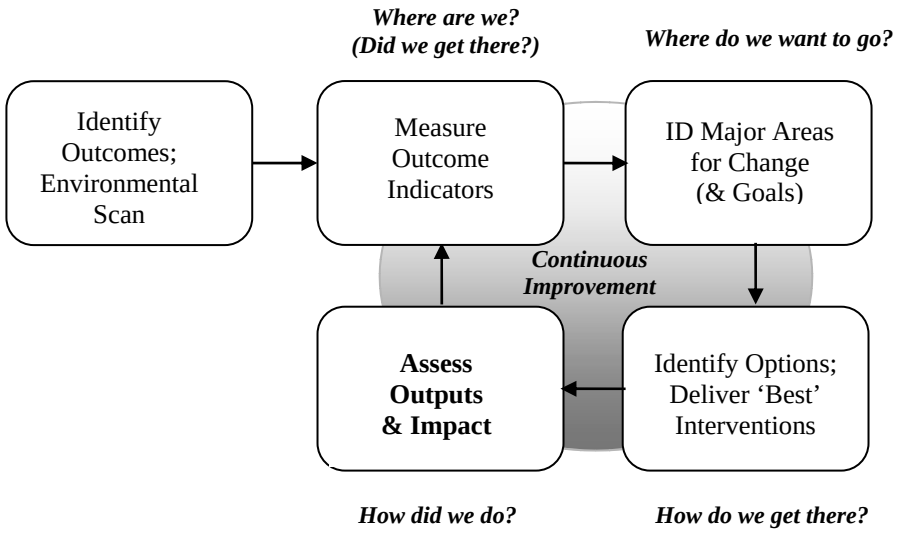
"Even relatively small increments of knowledge about how and why programmes work or fail cannot help but improve programme effectiveness."

Carol Weiss (2000)[1]

Purpose Statement

1 This Building Block outlines some challenges of, and options for, assessing the effectiveness of major interventions. This is the fourth of five papers on building evidence about outcomes into management systems.

2 Impact assessment provides feedback about performance, and closes the continuous improvement cycle (below). This paper discusses the need to assess the impact of interventions, provides an overview of some assessment methods, and discusses the use of impact information in decision-making. The advice is written for management and teams tasked with building results-based governance systems.



Management Overview

3 Some interventions do not work – or have unintended consequences – despite our best efforts.

4 Impact assessment helps departments make good decisions on where, when and how to intervene. Feedback allows us to prioritise interventions that deliver most benefit to the public, for the taxes we take. Impact assessment can be used to defend funding by demonstrating effectiveness to Parliament and the public.

5 Impact assessment works by comparing the outcomes of interventions with estimates of what would have happened without them.[2] Results are used in performance reporting, budget setting processes, as triggers for reviews, and as inputs to in depth reviews of the output mix or alternative service delivery options.

[1] In Rogers *et al*, eds. Program Theory Evaluation: Practice, Promise, and Problems, *New Directions in Evaluation*, N°. 87
[2] After the definition of Impact Evaluation, p.5, US General Accounting Office, GAO/GGD-98-26, April 1998.

6 Impact is assessed using the most robust approach that is feasible. Mixed methods are often used. Use the methods and information — qualitative, quantitative, objective or subjective — that best estimate outcomes. Where feasible, using standardised methods will help you to compare the impact of alternative interventions.

7 When the cost of assessing impact across a number of interventions is similar, first priority should given to assessing interventions that:

- represent a higher risk to the agency and its clients (interventions may not work as planned);
- consume significant resources (in total or per case);
- are expected to have a big impact at modest cost; and/or
- are being piloted and have significant capacity for growth.

8 We should at least try to show our interventions do not cause harm.[3] But impact assessments are hard to produce, and cannot always be produced. The Government often works in environments where it is difficult to link cause and effect. Impact may not be evident for a long time, and there are practical and ethical challenges in assessing the impact of what we do.

9 When impact cannot be assessed directly, monitoring approaches based on intervention logic provide timely information on whether our interventions work as intended (Building Block 5). At a conceptual level, intervention logic can be used to identify outputs that have a poor fit with the outcomes sought. We can also monitor the quantity and quality of our outputs, whether we get outputs to the right people (coverage) and near-term results. Monitoring may not show impact but it can build confidence that interventions work, and show where others need to be improved.

10 Positive information on impact can be used to protect or increase funds for effective interventions. Negative results – from impact assessment or monitoring – should trigger corrective action such as service improvements or budget reductions. Monitoring information can give timely feedback on where and how we need to improve.

11 But take care how you use information. Give new interventions time to mature, and be cautious if you extrapolate positive results for an intervention from one group to another group. Results can vary markedly with service delivery, between different client groups, between providers and over time.

12 Be vigilant for changes in outcomes you are not monitoring.

Making Impact Assessment work

13 Impact assessment is simply a process for confirming that our interventions produce positive outcomes. A variety of evaluative activities may be used depending on the circumstances. Impact assessment may involve simple analysis of intermediate outcomes during implementation, full evaluations (including formative, process, and impact), experimental designs, or low cost monitoring approaches that confirm interventions still work well.

[3] The term interventions includes legislation, policies, transfers, programmes and service delivery arrangements

Exhibit 1. Some approaches to monitoring impact assessment work

Government agencies in New Zealand improve impact and value-for-money in various ways, for example:

- Where impact assessments are not yet available, the Department of Conservation uses intervention logic to eliminate outputs that have a poor conceptual fit with the outcomes sought. If the 'vital few' outcomes are not measurable, strong conceptual approaches may need to be developed.
- The Land Transport Safety Authority combines accident data and impact information to estimate how new intervention approaches should reduce the social cost of accidents. State Indicators are monitored using statistical process control to confirm overall results were positive.
- A number of organisations in the health sector assess the impact and cost-effectiveness of medical treatments using information from published studies. They include the New Zealand Guidelines Group (NZGG) and Pharmac. This approach is economical where delivery factors are controlled at output level, and external effects are weak. The NZGG involves professionals and consumer groups in developing and implementing guidelines. (See http://www.womens-health.org.nz/guidelines.htm for the benefits of involving consumers.)

14 Impact assessment serves two main functions. First, it helps us make individual interventions work as well as they can. Second, it helps us shift resources to the interventions that improve outcomes most cost-effectively. Impact assessments cannot always be produced, but agencies that routinely use impact assessment tell us:

a *The intervention mix must work as a package.* We want to know whether individual interventions work, and that the 'package' of interventions delivered to individuals works overall. We also want to find combinations of interventions that can address different problems experienced by different groups.

b '*Evidence trumps opinion every time'* (New Zealand Guidelines Group). Decision makers at all levels must commit to identifying 'what works' based on the best evidence they an get. Expand good services. Redesign or downsize services that under-perform. But avoid treating staff harshly – poor incentives create resistance.

c *Delivery matters.* High service standards ensure the benefits assessed today are delivered into the future. Engage senior managers, end users and key stakeholders. Early engagement helps build users' confidence in the results, and helps manage expectations. Assessments that are trusted will be used to make decisions. Consult with groups receiving interventions, as well as staff. All forms of assessment have cultural bias. Impact may need to be assessed from multiple perspectives (the NZGG offers helpful guidelines on this).

d *Use good methods*. Impact assessment is hard, and it can be costly. There is no 'one right way'. The 'best' methods vary from intervention-to-intervention, and how and when we will need to use results. Our approach is shaped by many factors, e.g. culture, ethics, external factors, data limitations and affordability.

e *Assess mature interventions*. Let interventions mature and have the opportunity to work, before assessing impact. In the short-term, ensure interventions are designed and delivered well, and reach the right people.

What Interventions Should Be Assessed, and For Which Groups?

15 Impact assessment helps us make better resource decisions by answering three basic questions:

- Does an intervention or group of interventions improve outcomes as expected?
- Where, when and with whom do our established interventions work best?

- Which interventions are likely to be most cost-effective?

16 These questions provide a touchstone for identifying where impact should be assessed first. When the cost of assessing different interventions is similar, high priority is usually given to assessing interventions that:

- represent a higher risk to the agency and its clients (interventions do not always work as planned);
- consume significant resources, either in total or per intervention;
- are expected to have a big impact at modest cost; and/or
- are being piloted and have significant capacity for growth.

Priorities then shift to assessing interventions that are less risky, cheaper, or have less capacity for growth.

17 We also want to know who benefits most. Comparing the impact of an intervention across different groups can help us improve our criteria for who should be eligible for a service or intervention, or who should be encouraged to use a service or intervention. Where services, for instance, are delivered in large quantities it may be possible to assess impact separately for major groups. Conversely, comparing impact across sites helps identify and spread best practice (see the Pathfinder supporting paper "Benchmarking using outcome information").

18 Impact assessments only show that an intervention worked for the group it was 'tested on'. Having assessed a selective intervention, it is best to use clear allocation criteria to ensure it continues to be delivered to a group that will benefit. When services, policies or programmes are extended to groups that differ from the group originally assessed, a new assessment may be needed.

19 Re-assess impact periodically when results are affected by factors outside the intervention (non-treatment variables). The impact of social interventions can also vary with changes in targeting, group composition and delivery. The feasibility of repeat assessments depends on budget, evaluation priorities and evaluation costs.

Exhibit 2. Implementing Straight Thinking

Impact can change with implementation, between delivery sites, and across target groups. For example:

- When Straight Thinking (ST) was first piloted in Canada it reduced re-offending rates by 15 percent. When ST was rolled out nationally, this fell to only two percent. Delivery factors now get a lot more attention.
- When ST was implemented in the United Kingdom, impact varied five-fold across target groups of different re-offending risk (Findings 161, Home Office, UK). Working from offender histories, the Department of Corrections (NZ) suggests that delivering ST to the right group improves cost-effectiveness up to 14 times.

How Much Should We Spend on Impact Assessment?

20 The quick answer is that we should spend nothing, unless we are willing to change the way that we work. Spending on impact assessment may be limited to a few percent of the budget in areas where:

- impact assessment has previously confirmed that our major interventions are effective;
- service specifications and delivery standards have not changed; and
- external factors are not expected to influence impact.

21 Where impact has previously been determined, simple monitoring of the 'results chain' – how well outputs are delivered, coverage, and near-term results — can maintain confidence in the continued effectiveness of interventions. In-depth assessments can then be triggered if performance problems become evident.

22 We may need to spend significantly more on impact assessment when unproven interventions dominate our outputs, and where we need to assess the impact of major new interventions or pilot programmes.

How Can Impact Be Assessed?

23 The hardest decisions in assessing impact revolve around choosing the method.

24 Good methods produce accurate, affordable and timely estimates of impact that are strongly attributable to our interventions. But selection is influenced by other important factors. Good methods:

- are aligned to the outcome of interest – so a mix of qualitative and quantitative tools is usually needed;
- produce performance information in an ethical and culturally sensitive manner; and
- allow us to factor assessment errors into decision-making.

25 Key steps in identifying the most appropriate method include:

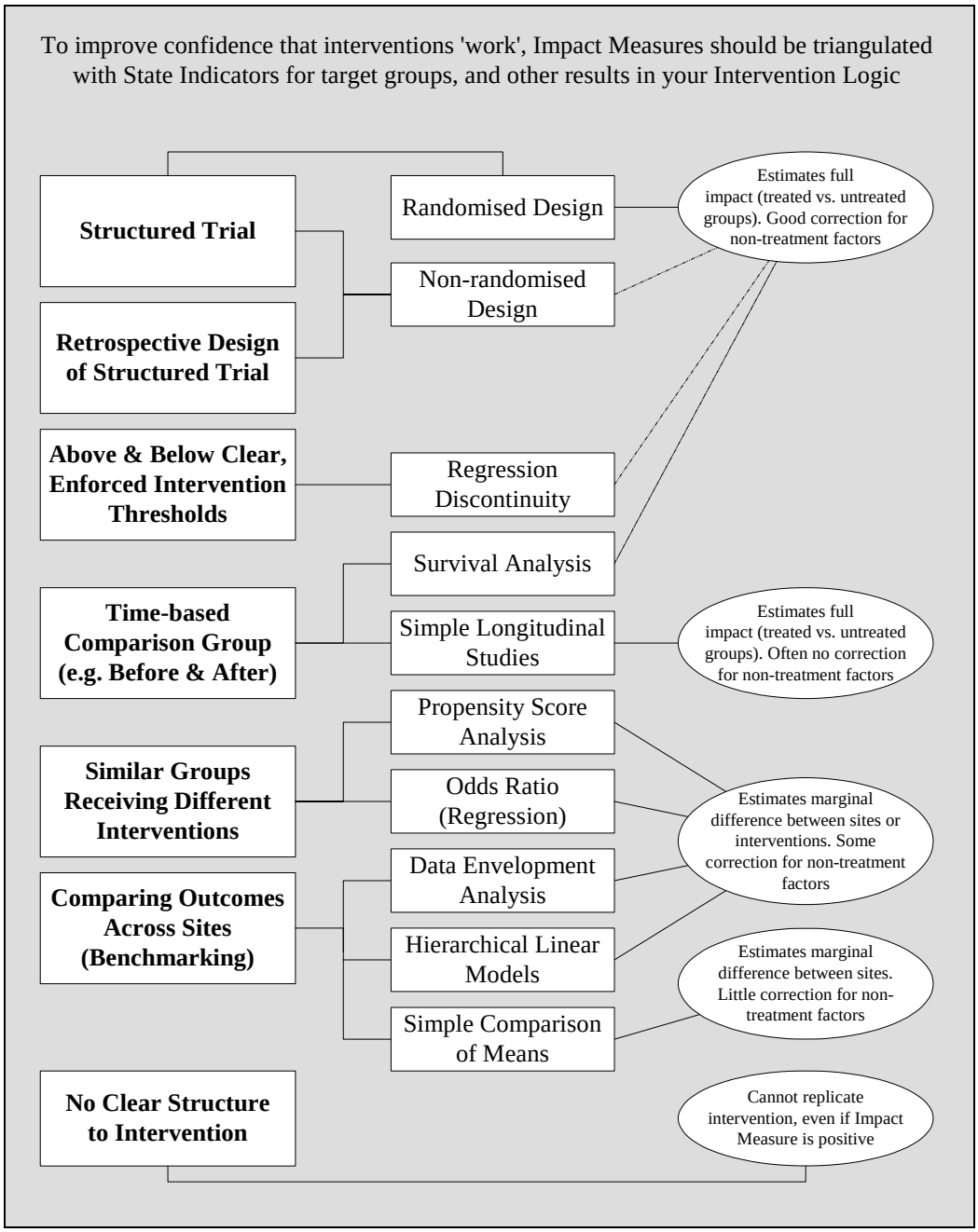
- specifying the outcome information needed, including intermediate outcomes or proxy measures;
- identifying interventions for which reliable impact assessments are already available[4];
- listing interventions to be assessed, including other interventions to be assessed in comparable terms;
- identifying information needs and costs, including contextual and supplementary information;
- assessing whether and how comparison groups can be set up in an ethical way (Exhibit 3); and
- determining how external, group composition and programme factors will be allowed for.

26 Intermediate outcomes, immediate effects and proxy measures are often used to assess likely outcomes where end outcomes cannot be measured in the near-term. For example, in health care we monitor the proportion of adults who smoke cigarettes to assess the impact of anti-smoking interventions, though the end outcomes sought are increases in life expectancy and reductions in the incidence and impact of cardio-vascular diseases. Other interventions, e.g. structural and procedural changes in the health and disability sector, are difficult to quantify directly, and impact is assessed using qualitative tools.

27 Ideally, we want to identify interventions that improve outcomes the most – without unintended effects. To do this we compare the benefits of alternate interventions. Full measures of impact with good compensation for non-treatment factors help us compare interventions assessed at different times. To produce these measures we want to find an 'untreated' group[5] and want descriptive information on the composition of 'treated' and 'untreated' groups. When an 'untreated' group does not exist, e.g. due to ethical or statutory entitlements, we can compare outcomes across similar groups receiving a different mix of interventions.[6]

[4] Including, where non-treatment factors have little influence, impact assessments from the international literature.
[5] This is a misnomer, as 'untreated' groups often receive standard services but do not get the special intervention being tested. Wait-list comparison and regression discontinuity methods allow us to treat people in need as best we can.
[6] e.g. comparing a pilot programme such as Multi Systemic Therapy with our current 'best practice'.

Exhibit 3. Comparison Group Options, Illustrated Using Selected Quantitative Methods



28 Comparison group methods are often used to measure the impact of elective interventions during pilot studies. But they require expertise and advance planning to conduct well. When ethics, sample size and other factors make it hard to create separate comparison groups, different methods must be used. Separate comparison groups have, however, been used in the following situations:

- insufficient funding exists to treat all eligible cases (multiple options exist);
- logistic factors prevent all eligible cases receiving treatment (use wait-list comparisons, etc);
- eligibility is determined by enforced entry requirements (use regression discontinuity);
- impact is known to vary with unidentified differences in delivery (multiple options exist); and
- an new intervention is delivered from a specific date (use time series analysis).

29 Many methods exist for assessing the integrity and likely impact of individual interventions. These methods should be used routinely, especially when direct measures of impact cannot be produced. Intervention logic and other forms of structured reviews are commonly used in New Zealand and Australia.

In situations when 'untreated' groups are difficult to find, we can compare how different groups responded to the same intervention, looking for clues about the effect of context and what drove success. In many social policy situations, techniques like comparative analysis are used. Here cases where there are similarities between key relevant socio-economic variables are carefully compared to understand the differences between them.

30 Intervention logic helps us identify the chain of results we expect to see if an intervention is working well. Monitoring these results builds confidence that interventions work (or at least work as they were designed to). The 'results chain' includes the quantity, quality and coverage of outputs, near-term results and intermediate and end outcomes. When end outcomes cannot be measured, we focus on showing that outputs are delivered in the right volume, and with the quality and coverage we said were needed. We then need to show that near-term results and intermediate outcomes are consistent with what needs to happen for end outcomes to improve. Simple trends in state indicators (e.g. for intermediate outcomes) help build confidence that interventions work.

31 Impact assessment and monitoring based on intervention logic can be a powerful combination. Monitoring gives timely feedback on how well specific aspects of the intervention are working, allowing course correction. Impact assessment gives slower feedback about effectiveness and value-for-money. But we can seldom get all the performance information we would like. Management decisions are based on the best evidence we can get.

32 Context, ethics and information constraints all shape the assessment method we choose. Measuring impact well requires experience. We should emulate best practice in our field. But we should also discuss best practice with experts in evaluation and statistics, and explore new approaches with them.

33 Detailed discussion of impact assessment frameworks is beyond the scope of this work. The following additional points should, however, be considered as we firm up our impact assessment approach[7]:

- impact assessment information must be structured to inform the decisions we have to make;
- to make timely decisions, we want to assess impact as soon as possible after we intervene. This is usually after benefits plateau. End outcomes can sometimes only be assessed years after an intervention ends (e.g. early education). Intermediate outcomes or proxies can be assessed when feedback is slow;
- best practice assesses outcomes across all people (or cases) receiving interventions. Non-completion of an intervention can signal problems or undue risks. Excluding people who drop out or are rejected can inflate impact assessments, and make us think an intervention was more successful than it was; and
- benefits often vary across groups – we may need to assess impact on different groups[8].

[7] Other criteria for measuring impact are implicit in scoring criteria used in meta-analysis and evidence-based approaches to identifying interventions that work, e.g.: http://www.ncjrs.org/works/appendix.htm. The Campbell and Cochrane Collaborations also list criteria supporting decisions on 'what works' in social policy and medical practice, respectively.

[8] Re-offending is often measured as the percentage of prisoners convicted of any offence within a fixed period after release – this does not allow for marked variation in the number and severity of offences between target groups. The social and economic benefits from successful hip replacement or heart surgery vary markedly with the level of debility suffered by different patients.

Generating Momentum

34 Action is often required now. New services can still be rolled out, even when results are uncertain. But in deciding to act we must balance the ethics and risks of implementing changes where results are uncertain, with the ethics and risks of doing nothing and the benefits we lose from services we discontinue. To help manage implementation risks, new services should generally be piloted – and evaluated - before national rollout.

35 If results are uncertain, it is often unethical to intervene without doing our best to prove the intervention works and managing major risks. Early evaluation results may focus on delivery, outputs and immediate effects. But in the longer term we try to assess intermediate and end outcomes

What Data and Information is Needed?

36 If we already produce state indicators for an outcome, we may have some outcome information needed to produce basic impact measures. But extra information is often needed to check for unintended results, create comparison groups, find out who benefits most, and allow for interactions between interventions. So we list the information needed and confirm data is available. Extra information may be needed to:

- create 'recipient' and 'comparison' groups of similar composition[9];
- identify all the interventions someone received (with start and completion dates);
- confirm 'recipient' and 'comparison' groups are similar except for the interventions received ;
- compare the impacts of an interventions across different recipient groups; and
- provide contextual information and test assumptions in the intervention logic.

37 To compare results across interventions and over time, we must put standards in place to ensure information is recorded consistently. Outcomes and classification data must be defined clearly. When qualitative measures are used to assess subjective experience (e.g. safety or disability), precision and repeatability can be enhanced asking respondents to scale qualitative measures against graduated statements of how experiences affect them.

38 Information produced from qualitative research is often required to understand and interpret the impacts of interventions. Exploratory research is often used when the factors affecting an intervention's success are not well understood. The Department of Labour uses several qualitative methods:

- case studies to reflect aspects of the matter under study or theory being tested;
- focus groups to get feedback on an issue or idea from different perspectives;
- surveys to collect data on perceptions across a population (or part of it);
- in depth interviews with key informants to explore issues in greater depth; and
- to provide quality assurance, they work closely with researchers doing other types of research.

Assessment, Reporting and Decision Making

39 *We never have perfect information.* The challenge is to make good decisions using the information available now, while gathering rich information to support decisions we have to make in the future. Building the evidence base is a bit like making a quilt. We start off with bits of information that do not fit together particularly well. But over time we gather extra bits, and shape the information base to support the decisions that must be made.

[9] To create time series comparisons, we may only need information confirming the intervention was received, and any extra information we require to compare results for different groups of recipients.

40 *Decide what works best, for whom, when*. Impact assessment reports must help decision makers agree on:

- the extent to which the agency's intervention logic and output mix is robust;
- the quantity, quality and coverage of outputs that will be funded into the future;
- which interventions need to be improved; and
- whether new interventions are required in some areas.

41 *Impact assessments provide only one input – but a key input* – into the above decisions. Information about the quantity, quality and coverage[10] of current outputs helps us fine-tune the current output mix. But major changes in intervention logic or output mix are most likely to be triggered by a shortfall in results – either as near term results, intermediate outcomes or end outcomes. Positive assessments tell us what to protect or grow. Negative feedback prompts decisions on our overall approach, and what should be improved or reduced in size.

> "Agencies' use of inconsistent definitions for their programs' measures could hamper decision-makers' use of data collected from those measures in planning, comparing performance, and reporting on performance achieved."
>
> *Implementing results-based management: lessons from the literature.* Office of the Auditor General of Canada, 2001

42 *Move from analysis to decisions.* Ultimately, decision makers need to agree actions and budgets. Reports should identify output options, discuss evidence of likely impacts and costs, identify groups most likely to respond well to interventions, and recommend the outputs that should be delivered into the future. It is equally important to decide what we should stop doing. This frees up resources for all the good things we want to do.

43 *Be alert for unintended outcomes*. These may include unmeasured side effects (good or bad), variation in impact between groups or entities, and failing to get interventions to target groups (coverage). Unplanned, negative outcomes are reasonably common. Do not ignore evidence that conflicts with hopes and expectations.

44 *Avoid premature decisions, but make timely decisions.* First results for immature interventions may not be reliable. Be cautious in closing pilots and new interventions, but try to avoid expanding pilots until benefits and side effects are known. Early termination may be a good response to interventions with low benefit or harmful side effects. In many cases, other information may suggest interventions could be fine tuned or left to mature.

45 *Persevere.* Even partial gains in knowledge boost our understanding of how and why programmes work or fail. Take the steps required to improve our performance information over time, and ensure that the evidence we can collect is used to shape our strategy, Statements of Intent, Output Plans, interventions and budgets.

[10] Coverage matters. A perfect medical treatment that cures everyone but reaches only 50% of the sick is still only 50% effective in improving outcome indicators. Conversely, a great intervention delivered to the wrong people will deliver little value.