

## ADDRESSING ATTRIBUTION THROUGH CONTRIBUTION ANALYSIS: USING PERFORMANCE MEASURES SENSIBLY

John Mayne  
Office of the Auditor General of Canada  
Ottawa, ON

**Abstract:** The changing culture of public administration involves accountability for results and outcomes. This article suggests that performance measurement can address such attribution questions. Contribution analysis has a major role to play in helping managers, researchers, and policymakers to arrive at conclusions about the contribution their program has made to particular outcomes. The article describes the steps necessary to produce a credible contribution story.

**Résumé:** Le changement dans la culture de l'administration publique engage la responsabilité des résultats et des conséquences. L'article suggère que mesurer le rendement peut répondre à ces questions d'attribution. L'analyse de la contribution est très important pour aider les gestionnaires, les chercheurs et les décideurs à parvenir à des conclusions sur la contribution de leur programme pour arriver à des résultats particuliers. L'article décrit les étapes nécessaires pour produire une histoire de contribution crédible.

■ A significant element of public-sector reform in many jurisdictions is the move away from a management regime focused on rules and procedures toward an approach that pays greater attention to the results being sought with taxpayers' dollars. *Managing for results*, *results-based management*, and *performance management* have become common terms in public-sector reform discussions (Auditor General of Canada, 1997; Organisation for Economic Co-operation and Development [OECD], 1997; President of the Treasury Board, 1997). The aim is to change the culture of public administration from one that is rules focused to one focusing on the results that matter to citizens. This approach is characterized by measuring progress toward the results that are sought, having the flexibility to adjust operations to better meet these expectations, and reporting on the outcomes accomplished. Some jurisdictions have legislated this approach to public administration.

In many cases, progress has been made in moving in this direction. Nevertheless, the challenges of managing for results have been and remain significant, in particular the difficulty in the public sector of measuring outcomes and of establishing the links to the activities of the program in a cost-effective manner. Some of these measurement problems are discussed below. There is an additional, related problem that has not received enough attention: the need to rethink how we deal with accountability in this new management paradigm.

## ACCOUNTABILITY FOR OUTCOMES<sup>1</sup>

In the past, accountability for the processes followed, the inputs used, and perhaps the outputs produced was most likely to be the arena in which public servants worked. This focus was consistent with the more traditional view of accountability: emphasizing what could be controlled and assigning blame when things go wrong. If the expected process was not followed, improper inputs were used, or outputs were not delivered, then the responsible person could be identified and appropriate action taken, as one ought to be in control of the processes, the inputs, and the outputs. Given this paradigm, public servants often were reluctant to accept accountability for results beyond outputs, that is, outcomes over which one does not have full control. Being accountable for outputs has been much more acceptable to public servants than being accountable for outcomes. And in these cases, establishing the links between activities and outputs — attribution — is not a significant issue: it is clear that the program produced the outputs.

In the case of managing for results, and in particular outcomes, the degree of administrative control and scope for influence a manager has over the outcomes sought will vary considerably in different situations. In some cases, the program manager in question is the main player and has a quite significant degree of control over the outcomes. In other cases, the manager might be only one of several players trying, with the resources and authorities available, to influence the achievement of the intended outcomes. Effective accountability implies that managers understand these considerations and have the means to deal with these more complex situations.

If the expected outcomes have not been accomplished, there may be several reasons, only one of which may be that the “responsible” manager has not done a good job. The manager might indeed have done all that could be expected, but the results were not achieved

due to circumstances beyond his or her influence. To encourage and support managing for results, we need a new view of accountability that acknowledges this more complex management world (Hatry, 1997). Attribution here is a real problem.

Accountability for results or outcomes<sup>2</sup> asks if you have done everything possible with your authorities and resources to effect the achievement of the intended results, and if you have learned from past experience what works and does not work. Accounting for results of this kind means demonstrating that you have made a difference; that through your actions and efforts you have contributed to the results achieved. Finding credible ways to demonstrate this is essential if the move toward managing for results is to succeed.

### THE PROBLEM OF ATTRIBUTION

Government programs are intended to produce certain outcomes: more jobs, a healthier public, better living conditions, and so on. Effective programs are those that make a difference in meeting these kinds of objectives — they contribute to the intended outcomes that citizens value. In trying to measure the performance of a program, we face two problems. We can often — although frequently not without some difficulty — measure whether or not these outcomes are actually occurring. The more difficult question is usually determining just what contribution the specific program in question made to the outcome. How much of the success (or failure) can we attribute to the program? What has been the contribution made by the program? What influence has it had?

Despite the measurement difficulty, attribution is a problem that cannot be ignored when trying to assess the performance of government programs. Without an answer to this question, little can be said about the worth of the program, nor can advice be provided about future directions. Perhaps even without the program the observed changes in outcomes would have occurred, or would have occurred at a lower level or later. In most cases, there are many other factors at play in addition to the impact of the program's activities. Such things as other government actions or programs, economic factors, social trends, and the like can all have an effect on outcomes. Managers, the government, and taxpayers would like to know the program's contribution to assess the value of continuing with the program in its current form. Unless we can get some handle on this measurement problem, accountability for results will never take hold.

The question is, how can we demonstrate that a program is making a difference?

Policy and program evaluation is one measurement discipline that tries to provide answers to this attribution question (Freeman & Rossi, 1993; Hudson, Mayne, & Thomlinson, 1992; Wholey, Hatry, & Newcomer, 1994). Traditionally, it uses some form of controlled comparison to estimate what happens with the program in place versus what would happen without it. Extensive social science research methods have been designed with this problem of attribution in mind. And an evaluation study probably remains the best way to address this problem, if one has the time, money, and expertise.

## THE CASE OF PERFORMANCE MEASUREMENT

Performance measurement is extensively and increasingly used to measure the performance of government programs (Mayne & Zapico-Goni, 1997). In contrast with evaluation, which usually undertakes special one-time measures and extensive analysis of the data gathered, performance measurement is characterized by regular and often more straightforward measurement of aspects of a program's performance. Performance indicators are used to track performance and as feedback information to managers and staff. They can form the basis for reports on what has been achieved by the program.

Performance measurement is often aimed at the very first level of impacts of a program, namely measuring the specific outputs (goods and services) provided by the program personnel (see Figure 1). In these cases, the question of attribution is not likely to be a problem because there is an evident direct link between what the staff are doing and their immediate products. Increasingly, however, organizations are trying to measure or track the subsequent impacts of these services and products, the intermediate or even more end outcomes they are trying to accomplish. The attribution issue quickly surfaces. In the absence of a thorough evaluation study, what can be done?

It is possible to structure a performance measurement system to try directly to get a measure of attribution. One could construct time series data by modifying the program over time and tracking the resulting changes. Or, in addition to measuring the impacts on those who are receiving the program, one could also measure the changes occurring in a similar comparison group that does not receive the

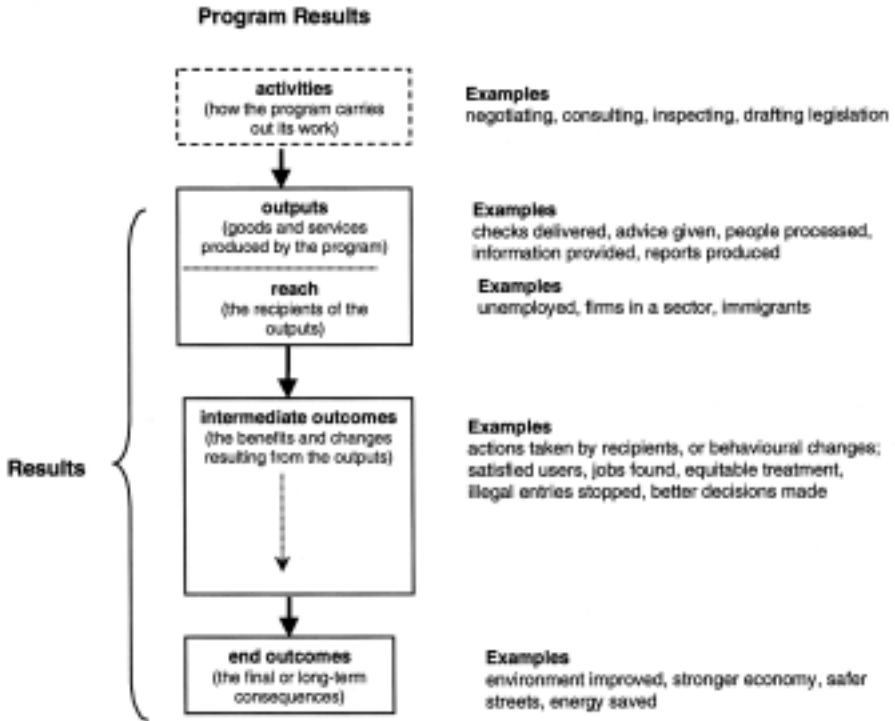
program. These approaches become *de facto* evaluations, using some form of quasi-experimental design.

Although possible, this carefully constructed and often expensive measurement strategy is not usually associated with most performance measurement approaches. In the absence of an evaluation study, what can one do in the case of a “normal” or typical performance measurement or monitoring system to get a handle on the attribution issue? This is the question addressed in this work.

RECOGNIZING THE LIMITS OF MEASUREMENT

First, we must recognize that determining definitively the extent to which a government program contributes to an particular outcome

**Figure 1**  
**Types of Program Results**



is usually not possible, even with a carefully designed evaluation study. We might be able to provide considerable evidence regarding a program's impacts and might be able to significantly increase our understanding of how a program is influencing a certain outcome, but in most cases of any complexity, there will not be a 100% guarantee. Rather, we need to talk of reducing our uncertainty about the contribution of the program. From a state of not really knowing anything about how a program is influencing a desired outcome, we might conclude with reasonable confidence that the program is indeed having an attributable impact; that it is indeed making a difference. We might also be able to provide a reasonable estimate of the magnitude of the impact.

Thus, we may need to rethink what measurement can usefully mean. Measurement in the public sector is less about precision and more about increasing understanding and knowledge. It is about increasing what we know about what works in an area and thereby reducing uncertainty (Auditor General of Canada, 1996, p. 21). This view of measurement implies that we can almost always measure things, and in particular the contribution a program is making. That is, we can almost always gather additional data and information that will increase our understanding about a program and its impacts, even if we cannot "prove" things in an absolute sense. We need to include softer and qualitative measurement tools within our concept of measurement in the public sector.

The limits of measurement mean that we need to accept some uncertainty about the measures of performance we are likely to have available in many cases. If you must know with a high degree of certainty just what a program's contribution is, then a well-designed evaluation is required. What I address in this work applies in cases where one is willing or is required to make do with less certainty, where the aim of measurement is to acquire some insight and develop some assurance that the program is actually having an impact. This, I suggest, is or ought to be the aim of performance measurement. A good overall measurement strategy would include both ongoing performance measurement and periodic evaluation.

## TWO USES OF PERFORMANCE MEASUREMENT: UNDERSTANDING AND REPORTING

We need to distinguish two uses that can be made of performance measurement information. First, performance information can be

used to better understand just what contribution a program is making. This is the management perspective, where one wants to use measurement to know more about if and how the program is making a difference; one is searching for knowledge. One wants to determine if the program is the appropriate policy tool to achieve the desired result. Here the question is how to use performance measurement as an investigative tool.

A second use of performance measurement is to explain or demonstrate the performance achieved by a program. In many jurisdictions, there is an increased focus and emphasis on reporting to elected officials and the public what has been achieved with the tax dollars spent and resources used. Performance measures frequently form the basis of such reporting. The question here is, how can performance measurement information be used to credibly report on what has been accomplished? In particular, how can it best be used to report on the contribution being made by a program?

We need to keep these two uses in mind as we consider how to deal with attribution using performance measures.

## APPROACHES TO ATTRIBUTION: CONTRIBUTION ANALYSIS

What is needed for both understanding and reporting is a specific analysis undertaken to provide information on the contribution of a program to the outcomes it is trying to influence. Coupled with the comments above about the nature of measurement in the public sector, the task at hand might be best described as, for reporting, trying to paint a credible picture about the attribution of a program. For understanding, the task is to glean as much insight as possible from performance measures about how well the operations of the program are working. In both cases, the key tool for the analysis is the results chain, which illustrates what is supposed to happen as a result of the activities undertaken and the outputs produced.

Too often, the measuring and particularly the reporting of performance through performance measurement systems completely ignores the attribution problem. The performance measured is either directly attributed to the program or attributed by implication, through the lack of any discussion or analysis of other factors at play. For anyone with even a little knowledge about the program and its environment, this kind of performance information will have little credibility. For managers, it provides no value-added information. In most cases,

any number of factors other than the program itself can be advanced to explain the observed outcome. The more obvious these other factors are, the less credible is the performance information. Discussing other factors may also provide insight into the program itself, how it operated, and its effects.

### Acknowledge the Problem

Thus, there is a need to acknowledge that there are other factors at play in addition to the program and that it is therefore usually not immediately clear what effect the program has had or is having in producing the outcome in question. Managers need to be realistic about the outcomes they are trying to influence if they want to gain new insight on how and if their activities are making a difference. For reporting, acknowledging the other factors at play is more honest and hence more credible than pretending they do not exist. As we will see below, there is more that can be done, but recognizing the other factors at play while still believing the program is making a contribution is a critical first step.

We suggest a number of steps, outlined in Figure 2, that can be used to address attribution through performance measurement. Collectively, these are elements of a *contribution analysis*.

Contribution analysis attempts to explore and perhaps demonstrate what Hendricks (1996) calls “plausible association”: whether “a reasonable person, knowing what has occurred in the program and that the intended outcomes actually occurred, agrees that the program contributed to those outcomes.”

#### Step 1: Develop the results chain

There is some logical reasoning behind the program that explains what it is supposed to be accomplishing and how. This logic or theory might be quite convincing or well established based on past experience. By developing the logical case, one can see what is supposed to or is believed to be happening. Constructing and presenting this theory of the program is a standard component of planning for an evaluation study (Wholey, 1983), where often a logic chart is used (Julian, Jones, & Devo, 1995).

More recently, the power of this approach is increasingly used in performance measurement where such terms as *outcome sequence charts*, *results chains*, and *visible indicator tree* (Meekings, 1995)



are being used to describe the same diagnostic tool. In addition to getting a handle on the attribution issue, these tools are proving invaluable in designing and implementing performance measurement systems. Further, by forcing program designers to be clear

## Figure 2 Contribution Analysis

---

**Step 1: Develop the results chain.** Develop the program theory model/program logic/results chain describing how the program is supposed to work. Identify as well the main external factors at play that might account for the outcomes observed. This program theory should lead to a plausible association between the activities of the program and the outcomes sought. Some links in the results chain will be fairly well understood or accepted. Others will be less well understood or subject to explanations other than that the program was the "cause." In this way you acknowledge that attribution is indeed a problem.

**Step 2: Assess the existing evidence on results.** The results chain should provide a good idea of which intended results (outputs, intermediate and end outcomes) could be measured. What evidence (information from performance measures and evaluations) is currently available about the occurrence of these various results? The links in the results chain also need to be assessed. Which are strong (good evidence available, strong logic, or wide acceptance) and which are weak (little evidence available, weak logic, or little agreement among stakeholders)?

**Step 3: Assess the alternative explanations.** Outcomes by definition are influenced not only by the action of the program but also by external factors — other programs, as well as social and economic factors. In addition to assessing the existing evidence on results, there is a need to explicitly consider the extent of influence these external factors might have. Evidence or logical argument might suggest that some have only a small influence and that others may have a more significant influence on the intended results.

**Step 4: Assemble the performance story.** With this information, you will be able to set out your performance story of why it is reasonable to assume that the actions of the program have contributed (in some fashion, which you may want to try and characterize) to the observed outcomes. How credible is the story? Do reasonable people agree with the story? Does the pattern of results observed validate the results chain? Where are the main weaknesses in the story? There always will be weaknesses. These point to where additional data or information would be useful.

If getting additional evidence is not possible (at least for now), then this is the most you can say about the extent to which the program has made a difference.

**Step 5: Seek out additional evidence.** To improve your performance story you will need additional evidence. This could involve information on both the extent of occurrence of specific results in the results chain and the strength of certain links in the chain. A number of strengthening techniques that you might be able to adopt are outlined in this work.

**Step 6: Revise and strengthen the performance story.** With the new evidence, you should be able to build a more credible story, one that a reasonable person will be more likely to agree with. It will probably not be foolproof, but will be stronger and more credible.

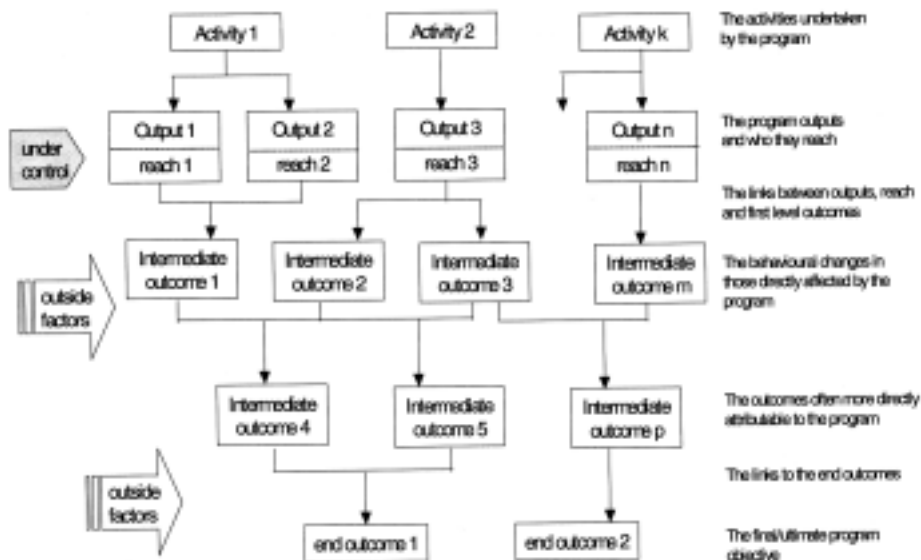
---

about the problems that programs are designed to address and how to address them, logic models or results chains encourage more precise program design.

A results chain or logic chart for a program tries to display on a page how the program is supposed to work — how, that is, the various outputs of the program are believed to produce a number of results that will lead to the intended end outcomes of the program. Results chains can also discuss unintended impacts that might occur and need to be watched for, as well as the key external factors influencing outcomes.

Figure 3 illustrates in a generic fashion what a detailed logic chart can look like; there are a variety of ways of presenting one. A logic chart illustrates the linkages between specific outputs, specific intermediate outcomes, and specific end outcomes. In other cases, it may suffice to present a less complicated picture of the program logic. Figure 4 illustrates this case for an environmental program. Logic charts explicitly include the idea of reach — whom the program is expected to reach (Montague, 1998) — and intermediate outcomes. This is because it is often at these levels that performance indica-

**Figure 3**  
**A Program Logic Chart or Results Chain**

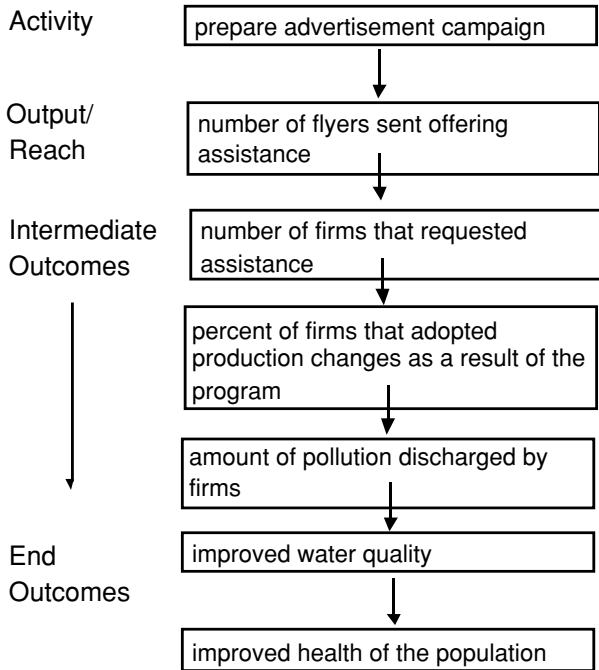


tors can do a good job of measuring; these are the levels in the results chain over which the program typically has most control. Further, evidence that the intended, more immediate outcomes have in fact occurred is a critical step in demonstrating the larger performance story. In this manner, the program can be shown to have had some effect.

Developing and using a logic chart has a number of benefits for program managers, such as developing consensus on what the program is trying to accomplish, developing an understanding of how it is believed to be working, clearly identifying the clients of the program, seeking and getting agreement on precisely what results are intended — the performance expectations — and identifying the key measures of performance. We are particularly interested in the additional benefits of identifying

**Figure 4**  
**Results Chain**

**Industrial Technical Assistance Program to Reduce Water Pollution**



*Source: World Bank, 1996.*

- the cause-effect relationships implicit in the program's theory,
- the outside factors at play, and
- areas where understanding about the impact of the program is weak.

Typically, some of the "links" between results in the logic chart are well known and have been established in past practice. There is less likely to be disagreement on their role in bringing about the intended outcomes. Other links may not be so well accepted and those suggest where further evidence (i.e., additional performance measures) might most fruitfully be sought. Any additional evidence one can gather to confirm such links will add to understanding of how the program is working and bolster the argument that the program is making a contribution. Similarly, if significant outside factors are identified as possibly having an effect on the intended outcome, then evidence to refute or determine the extent of influence of that effect will be useful in addressing the attribution question.<sup>3</sup>

In this way, managers can use the diagnostic tool of the results chain to better understand how they and others believe the program is working. They can design the program operations to fit these expectations. Through presenting and discussing the logic behind the program when reporting performance, one has laid out exactly what is being measured and what the major assumptions are concerning the contribution of the program. As a result, weaknesses in program assumptions are identified, suggesting where more evidence is needed. At a minimum, this kind of reporting allows one to know what challenges to the credibility of the performance measures used can be raised.

## Step 2: Assess the existing evidence on results

The results chain should point to which intended results (outputs, intermediate and end outcomes) could be and which should be measured. One needs to take a look at what evidence (information from performance measures and evaluations) is currently available about the occurrence of these various results. The links in the results chain also need to be assessed: which are well supported (good evidence available, strong logic, or wide acceptance) and which are weak (little evidence available, weak logic, or little agreement among stakeholders)? This assessment of the available information about the results chain will give you a good idea of where your perform-

ance story is strong and, more importantly, where it is weak and may be in need of additional data and information.

### Step 3: Explore and discuss plausible alternative explanations

The attribution problem arises when one believes or is trying to determine that a program has resulted in certain outcomes and there are alternative plausible explanations. That is, those who are skeptical that it really was the program's contribution that counted will point to other reasons for the observed outcome — for example, other related government programs, economic or social trends, or behavior unaffected by the program.

Dealing with these alternative explanations explicitly is often the best way of buttressing an argument in favor of the program's impact, the contribution it is having. This entails:

- identifying the most likely alternative explanations and the evidence associated with them;
- presenting whatever counter evidence or argument there may be and, where appropriate, discounting these alternative explanations; and
- presenting whatever evidence there is that the program is a more likely explanation.

Remember, the issue here is not a 0-1 situation. There likely are a number of factors in most cases that do in fact influence the outcome in question. No single factor, including the program, "causes" the outcome. Rather, the observed outcome is the result of a number of contributing factors. We want to understand better if the program is a significant factor in the occurrence of the outcome: that without the program in place, the outcome would probably not have occurred or would have occurred in a significantly different way.

If there is one or more plausible alternative explanations for the occurrence of the outcome, and if there is little evidence that counters these explanations, then you may have to conclude that you do not really know what the program's contribution has been and maybe (see below) suggest that an evaluation or further evidence is needed.

The kind of evidence that could be used to counter arguments for alternatives to the program depends on the program and its situation. But two generic types are available. First, there is a logic ar-

gument. One might refer to the theory behind the program — the results chain — and the kind of theory that would be needed to support claims for rival hypotheses. Supporting alternative explanations may involve assumptions more unlikely than those associated with the program. Second, one can bring actual evidence to bear concerning the alternative explanations, as discussed further on.

Addressing the attribution problem this way demonstrates that:

- you are aware of the complexity of the situation;
- you acknowledge and understand the other factors at play; and
- you are nevertheless concluding (assuming you are) that the most likely explanation for the observed outcome is that the program has made a significant contribution.

Others may try to demonstrate that some other factor than the program was the main factor in the chain of events that led to the outcome, but you are providing the best evidence you have for your case.

Unless you frankly discuss alternative explanations, your claim about the program's efficacy can be effectively challenged by simply pointing out the existence of alternative explanations.

#### Step 4: Assembling the Performance Story

With the information gathered, including that relating to alternative explanations, you will be able to set out your performance story of why it is reasonable to assume that the actions of the program have contributed to the observed outcomes. And you may want to try and characterize the nature and extent of the contribution claimed. How credible is the story? Do reasonable people agree with the story? Does the pattern of results observed validate the results chain? Where are the main weaknesses in the story?

There always will be weaknesses. These point to where additional data or information would be useful. Whether additional information is required depends on the particular situation. At present, what you have is what you have. In many cases, it may be best to conclude that at the moment there is not a great deal of certainty about the extent of the contribution being made by the program.

### Step 5: Seek out additional evidence

Once you have identified the weaknesses in the current performance story, there are many ways to better understand the results chain through gathering additional evidence. Building a performance story should be seen as a multi-year project — a journey — that is added to and modified as more knowledge is gained through monitoring and as the outside circumstances change. A number of approaches for strengthening performance stories are outlined below.

### Step 6: Revise and strengthen the performance story

As new evidence is gathered and assessed, as the results chain is populated with more information, a stronger story will emerge. The further evidence, of course, may not support earlier performance stories. The contribution being made by the program may not be as strong as originally expected, with the evidence pointing to the need to perhaps alter its activities.

## TECHNIQUES FOR STRENGTHENING YOUR PERFORMANCE STORY

In many cases, if a performance story is built only with the information currently available, numerous weaknesses in the evidence and logic will exist. The first iteration of a contribution analysis as outlined above will point to those weaknesses. A key aspect of the analysis is finding ways to strengthen the evidence about the contribution the program is making. Figure 5 outlines a number of strengthening techniques that can be employed, and that are discussed below.

### Refine the Results Chain

We may need to better understand the program in question, to dig deeper into what is supposed to occur as a result of its activities. A more detailed results chain is one approach, where the focus is on the specific behavioral changes resulting from the program's outputs that we can observe for those "reached" by the program. In order to bring about an outcome, programs have to change people's behavior.<sup>4</sup> The outputs of the program must be aimed at influencing the program's clients or target audience — the reach element of Figure 1 — to act in different ways so that the anticipated outcomes can occur.

Results chains often focus only on the sequence of events that are expected to occur, and thus may be at too aggregate a level to detect the specific behavioral changes that must occur as prerequisites of each of the events. By trying to identify and then document the

### **Figure 5** **Techniques for Strengthening a Contribution Analysis**

---

Any or all of these techniques could be adopted to strengthen the performance story. Which one is used depends on the specific situation.

**Refine the results chain.** It is often possible to develop a more detailed results chain, one that shows at a more micro level how one intermediate result leads to another. For example, many programs aim to change behaviour of the target group through communicating with them. Thus, intermediate steps are: ensuring the group in fact is aware of the messages communicated, understands them, and finds the messages reasonable. Each of these outcomes could be measured.

**Gather more results and/or linkages evidence.** It is rare that data on all steps in a results chain are gathered. This is usually not cost-effective. But this means that other data could be collected. In the case where the results chain is already reasonably detailed, efforts can be made to measure more of the intervening steps in the chain. If a more refined results chain has been developed, it will suggest additional, more refined data that could be collected to better understand key aspects of the results chain.

**Survey knowledgeable others.** There are likely a number of knowledgeable people who would have views on the extent to which the program was making a difference, either globally or at some intermediate point in the results chain. These might include the program recipients, partners involved in delivery, or outside experts in the area.

**Track program variations and their impact.** Where differences in the program can be identified, such as over time, between delivery locations, and among different target groups, gathering data on the corresponding differences in results patterns can provide valuable evidence in support of the overall performance story.

**Undertake case studies.** Examining one or a few specific case studies within the program can provide confirmatory evidence demonstrating that in these cases the program seems to be (or is) making a difference. On their own, a case study or two may not be too credible, but as part of a package of evidence in the context of a well-developed results chain, they do provide support for the reasonableness of the program theory.

**Identify other relevant research or evaluation.** In some program areas, there may be existing research or evaluation findings that provide supporting evidence for the results chain, or parts thereof.

**Use multiple lines of evidence.** Combining the evidence from several of the above approaches builds a more convincing, stronger case than any one line of reasoning.

**Undertake an evaluation.** Conducting a program evaluation may be the only way to gather further credible information on the contribution being made by a program. Indeed, effective measurement of results should be seen as a multi-year endeavor that includes both ongoing performance measures as well as occasional evaluation. Program evaluation should be part of a good *measurement strategy*.

---



changes in program target groups' attitudes, knowledge, perceptions, and decisions, which logically link to the outcomes being observed, we can often acquire a good understanding of the actual impact the program is having. If we can observe these short-term changes occurring, the logical case for the program's attribution can be enhanced. Furthermore, these are often some of the intermediate outcomes that can be measured more readily. As a result, it may be useful to set performance expectations and targets at this level, where there is a reasonable level of control (United States General Accounting Office [GAO], 1998).

Developing a logic model in this way takes us quite close to theory-driven approaches to evaluation (Chen, 1990, 1994; Weiss, 1997), where more detailed program logic models are developed to reconstruct the theory of the program; assess/test the credibility of the micro-steps in the theory (links in the program logic/story); and develop and/or confirm the results achieved by the program. Linking approaches for performance measurement with theory-driven evaluation is an area in need of more research.

Thus, managers trying either to better understand the effects of their programs or to report on performance can benefit from extending the analysis of logic charts to include consideration of the specific behavioral changes expected as a result of the program.

There is an additional way in which a logic model might usefully be refined. A good results chain for a program often illustrates the many aspects of performance that could be measured and reported. Considerable care is needed in selecting indicators of performance. It is important to use performance indicators that best discriminate or focus on the outcomes in question, yet often the indicators used relate only broadly to the circumstances of the program clients, the economy, or society as a whole. With a little more thought given to how the program operates (from the analysis of the results chain), the indicators can often be refined to more carefully focus on what specific benefits the program is intended to achieve. In particular, one can try and "refine the denominator" of the indicator.<sup>5</sup>

For example, many indicators are ratios, where the denominator qualifies the numerator. Consider a program designed to reduce air accidents by inspection of the air-worthiness of aircraft. An indicator might be the number of air accidents per air-mile flown. A better indicator would be the number of air accidents due to structural

failure per air-mile flown. But structural failures may occur regardless of inspections. Therefore, it may be better still to use two indicators: the number of air accidents per air-mile flown due to structural failure in aircraft inspected and the number of air accidents per air-mile flown due to structural failure in aircraft not inspected. By comparing structural failures in inspected and uninspected aircraft, one can estimate what inspection does to reduce the problems that inspection is designed to address. Questions of attribution still exist, but the more refined indicators reduce the problem and improve the chance of providing useful information on the contribution of the program.

### **Gather More Results and/or Linkages Evidence**

Performance measurement is about gathering evidence on the performance of a program. Our contribution analysis will have identified elements of the results chain where evidence is weak or lacking. We can set about bolstering the results chain with more or better evidence.

As suggested earlier, one might gather evidence concerning alternative explanations of the observed outcome. This will mean gathering data such as contextual and historical information about the plausibility of the alternative explanations. The data might be part of the routine performance measurement system, but more likely would be collected from time to time when analysis of the program's contribution is undertaken. Data collection might entail a review of the relevant literature, surveys, tracking of relevant external factors, field visits, or focus groups. The stronger the case that can be made, the stronger is the conclusion about the program's contribution.

Two sources of useful data are often overlooked. There is frequently considerable data available from *program files*, some of which might be useful to provide information on the contribution of the program. This type of existing data, which probably has been collected for other purposes, can often contain valuable information, particularly if used in conjunction with new data collected. In other cases, there may be useful *secondary analysis* available — studies that others have done in the program area that might clarify measurement and attribution issues. In still other cases, there may be *meta-analysis* that has been done — analysis that synthesizes a number of studies in an area.

## Survey Knowledgeable Others

As another source of evidence, one might try to gather information about the contribution of the program directly, most often through the use of *expert opinion*. In many program situations, there are persons outside the program who are seen as knowledgeable about the program area, the program's impacts, and the environment in which the program operates. A structured survey may be able to provide some evidence, albeit subjective in nature, of the extent to which the program is influencing an outcome. Surveying such individuals is often done to find out other information about the program, in which case adding questions on attribution is not very expensive. A focus group of experts may be another approach, and would allow some probing as to why views are held. In the absence of other, more costly data, this approach can be a relatively inexpensive way to increase confidence in the influence of the program.<sup>6</sup>

## Tracking Performance Variations and Their Impact

In cases where the program activities have varied over time, showing that outcomes have varied in a consistent manner with the variation in activities can strengthen the argument that the activities have indeed made a difference. In the simplest example, if an expected outcome has been observed after (and not before) the program activity has started up, this suggests the program is having an effect. In a more complicated case, if the outcome improves at sites (or at times) where the program has been implemented but not at others (such as a national program operating at many locations), the case for making a difference is even stronger. Hendricks (1996) identifies a number of such cases where, by tracking performance measures, we might show that:

- outcomes appeared at an appropriate time after our efforts began;
- outcomes appeared in different locations or with different people;
- outcomes faded when our efforts stopped;
- only those outcomes appeared that we should have affected;
- outcomes appeared only where or when we were active; and
- the biggest outcomes appeared where we did the most.

In some areas of programming, such as the impacts from research activities, there is likely to be a significant delay before the intended

outcomes occur, and the attribution picture portrayed through tracking performance over time will not be as evident. In these cases, one still needs to track outcomes over time to see if the intended outcomes have occurred, but demonstrating or understanding attribution is even more of a challenge. Some of the other approaches described in this article need to be used.

Tracking program variations is an approach that can come close to traditional evaluation methods for examining attribution.

### Undertake Case Studies

Use can often be made of *case study evidence* on a program's outcomes. In programs where there are specific cases, projects, or events, the evidence for attribution on one or two of these can be quite compelling; it can reveal the real nature of the program and also demonstrate, at least in these cases, that one can be fairly confident about the impact of the program's activities. In addition, case studies can also illustrate whether the program logic is indeed logical and reasonable (or not). This type of evidence can be quite persuasive, but appropriate cautions are a must, especially when the evidence is anecdotal. Case study and anecdotal evidence is most persuasive when illustrating a concrete case to complement other evidence that has been collected. On its own, however, it can be quite misleading, as it may merely be one of the few cases that appears to have worked among the vast majority that have not, as the U.S. GAO (1996) found in a review of "success stories" of the U.S. Department of the Environment. Further, readers are often tempted to generalize from anecdotal evidence, a practice that should be cautioned against. Nevertheless, if the context and limitations are made clear, there is often a useful role for individual case studies.

### Use Multiple Lines of Evidence

We have discussed a number of ways to deal with the attribution problem. We suggest that the more ways that are used in any one case, the more definitive information we will have on attribution. This is the "multiple lines of evidence" argument. Although no one piece of evidence may be very convincing, a larger set of different and complementary evidence can become quite convincing. Thus, in trying to reduce the uncertainty surrounding attribution, using as many lines of evidence as possible is a sensible, practical, and credible strategy.

## Undertake an Evaluation

In some cases, if the various lines of evidence point in different directions, there may be little one can say with enough credibility about the contribution of the program. If it is critical to have good information on attribution, then the best strategy may be to simply acknowledge that one does not know and suggest that an evaluation be carried out to address the attribution question. In most cases, however, if the program has indeed made a significant contribution, the various lines of evidence will confirm this.

## CONCLUSION

We have argued here that what is needed in dealing with attribution using performance measurement information is to explore the issue in a systematic way and, when reporting, to paint a credible picture of attribution to increase our knowledge about the contribution being made by the program. We need to accept the fact that in most cases what we are doing is measuring with the aim of reducing uncertainty about the contribution made, not proving the contribution made.

We suggest undertaking a *contribution analysis* that would examine and present the best case possible — a credible performance story — for attribution with the available evidence. In essence, we are suggesting *theory-driven performance measurement*, where, over time, using contribution analysis, a better understanding of just how the program is working is developed and used to improve future performance and report on past performance.

## A Credible Contribution Story

Using contribution analysis, a reasonable case that a program has indeed made a difference would entail:<sup>7</sup>

- providing a well-articulated presentation of the context of the program and its general aims, along with the strategies it is using to achieve those ends;
- presenting a plausible program theory leading to the overall aims (the logic of the program has not been disproven, i.e., there is little or no contradictory evidence and the underlying assumptions appear to remain valid);

- describing the activities and outputs produced by the program;
- highlighting the results of the contribution analysis indicating there is an association between what the program has done and the outcomes observed; and
- pointing out that the main alternative explanations for the outcomes' occurring, such as other related programs or external factors, have been ruled out, or clearly have had only a limited influence.

If all this is not enough, and there are too many gaps in the story, one ought to admit it and accept the need for an evaluation to provide better understanding of the contribution of the program.

Recognition of the problem and an understanding of the other factors at play will likely lead to additional data and information gathering. The result will be a better understanding of the program and how it is expected to work, and perhaps a redesign of the program to reflect this enhanced understanding. In addition, better performance information will provide for a more credible demonstration of the impacts of the program through performance measurement.

## NOTES

An earlier version of this article is posted on the Office of the Auditor General of Canada web site at <[http://oag-bvg.gc.ca/domino/other.nsf/html/99\\_e.html](http://oag-bvg.gc.ca/domino/other.nsf/html/99_e.html)>. Reproduced with the permission of the Minister of Public Works and Government Services, Ottawa, Ontario, Canada K1A 0S5, 2001.

- 1 This material is taken from a 1998 joint paper by the Office of the Auditor General and the Treasury Board Secretariat.
- 2 The terms *outcomes* and *results* are often used interchangeably. Strictly speaking, *results* includes outputs (see Figure 1) and hence is broader than outcomes. Nevertheless, much of the literature and some of the text here uses *results* to mean outcomes, when the intention is clear. If a reference is being made to outputs, then *outputs* will be used.
- 3 In the case of reporting, we are not suggesting that only evidence that bolsters the claim of program impact should be gathered or sought. Being able to say with some confidence that it is not known

- what contribution the program is making is also valuable knowledge. We are trying to gather through performance measurers as much evidence as is practical to understand the extent and nature of the contribution being made by the program and to support such a claim.
- 4 The ideas in this section were proposed by Steve Montague of the Performance Management Network, Ottawa.
  - 5 The term and the example were developed by Hugh McRoberts of the Office of the Auditor General.
  - 6 One caveat here is that if an individual expert has a vested interest in the program, then his or her views will need to be suitably discounted.
  - 7 Hendricks (1996) proposes a similar list.

## REFERENCES

- Auditor General of Canada. (1996). *Report to the House of Commons: Matters of special importance*. Ottawa: Author.
- Auditor General of Canada. (1997). *Report to the House of Commons, Chapter 11, Moving towards managing for results*. Ottawa: Author.
- Chen, H.T. (1990). *Theory-driven evaluation: A comprehensive perspective*. Thousand Oaks, CA: Sage.
- Chen, H.T. (1994). A panel on theory-driven evaluation and evaluation theories. *Evaluation Practice*, 15(1): 73–74.
- Freeman, H. E., & Rossi, P. (1993). *Evaluation: A systemic approach*. Thousand Oaks, CA: Sage.
- Hatry, H. (1997). We need a new concept of accountability. *The Public Manager*, 26(3), 37–38.
- Hendricks, M. (1996, November 6). *Performance monitoring: How to measure effectively the results of our efforts*. Paper presented at the American Evaluation Association Annual Conference, Atlanta, GA.
- Hudson, J., Mayne, J., & Thomlison, R. (Eds.). (1992). *Action-oriented evaluation: Canadian practices*. Toronto: Wall & Emerson.

- Julian, D.A., Jones, A., & Devo, D. (1995). Open systems evaluation and the logic model: Program planning and evaluation tools. *Evaluation and Program Planning*, 18(4), 333–341.
- Mayne, J., & Zapico-Goni, E. (Eds.). (1997). *Performance monitoring for public sector reform: Future directions from international experience*. New Brunswick, NJ: Transaction Publishers.
- Meekings, A. (1995). Unlocking the potential of performance measurement: A practical implementation guide. *Public Money & Management*, 15(4), 5–12.
- Montague, S. (1998). *Build reach into your logic chart*. <<http://www.pmn.net/contributions/reachlog.htm>>
- Office of the Auditor General of Canada & Treasury Board Secretariat. (1998). *Modernizing accountability practices in the public sector*. Ottawa: Author.
- Organisation for Economic Co-operation and Development. (1997). *In search of results: Performance management practices*. Paris: OECD.
- President of the Treasury Board. (1997). *Accounting for results: 1997*. Annual Report to Parliament. Ottawa: Treasury Board Secretariat.
- United States General Accounting Office. (1996). *Observations on DOE's success stories report*. Testimony before the Subcommittee on Energy and Environment, Committee on Science, House of Representatives. GAO/T-RCED-96-133. Washington, DC: Author.
- United States General Accounting Office (1998). *Managing for results: Measuring program results that are under limited federal control*. GAO/GGD-99-16. Washington, DC: Author.
- Weiss, C. (1997, Winter). Theory-based evaluation: Past, present and future. *New Directions for Evaluation*, 76, 41–55.
- Wholey, J.S. (1983). *Evaluation and effective public management*. Boston, MA: Little, Brown.
- Wholey, J., Hatry, H., & Newcomer, K. (1994). *Handbook of practical program evaluation*. San Francisco, CA: Jossey-Bass.
- World Bank. (1996). Colombia: Paving the way for a results-oriented public sector. Report no. 15300-CO. Washington, DC: Author.