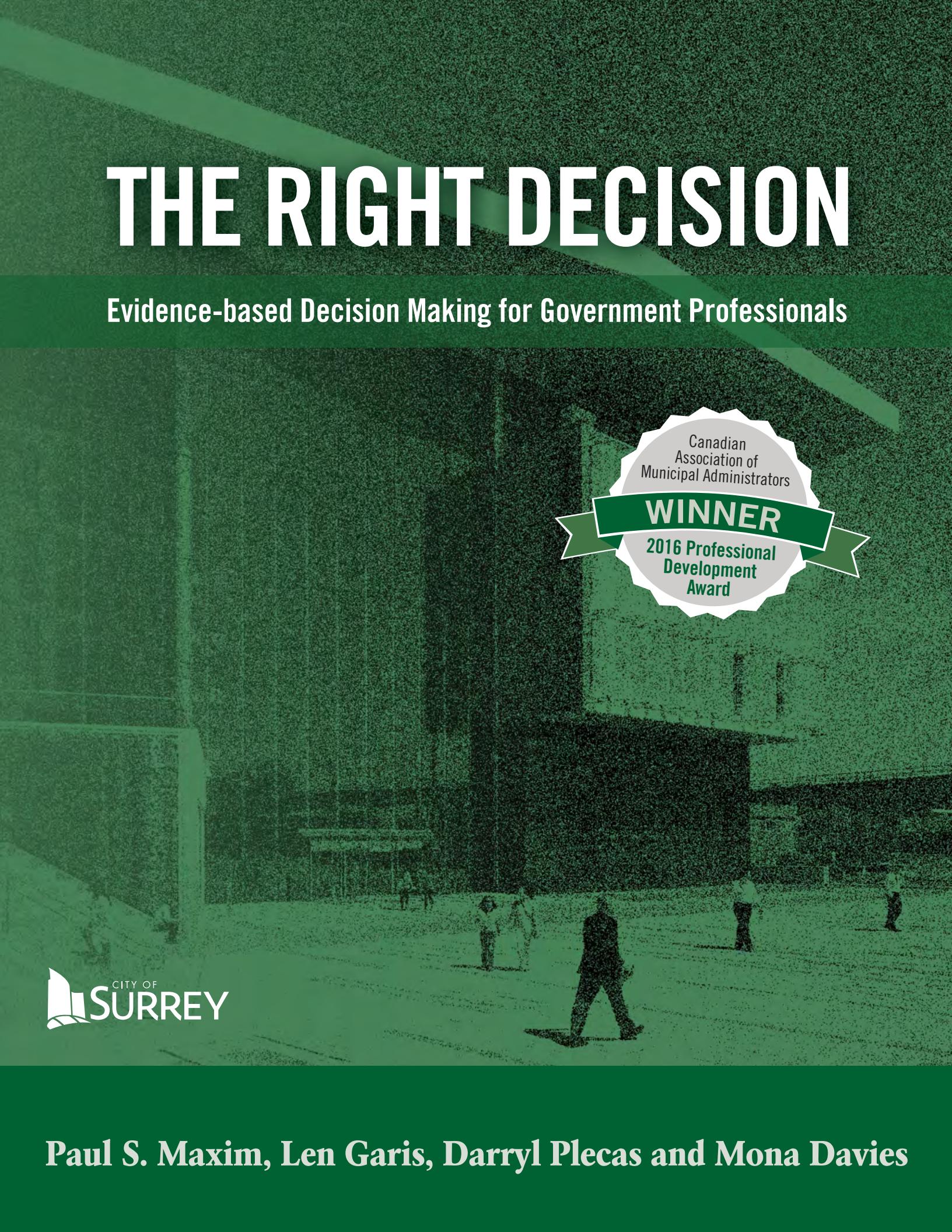


THE RIGHT DECISION

Evidence-based Decision Making for Government Professionals



Canadian
Association of
Municipal Administrators

WINNER
2016 Professional
Development
Award



Paul S. Maxim, Len Garis, Darryl Plecas and Mona Davies

For more information about this book or its authors, contact Len Garis at the University of the Fraser Valley, B.C., at len.garis@ufv.ca or 778-846-0851. For additional public safety publications, please visit <http://cjr.ufv.ca>

THE RIGHT DECISION

Evidence-based Decision Making
for Government Professionals

Paul S. Maxim, Len Garis, Darryl Plecas and Mona Davies

About the Authors

Paul Maxim, Professor, Wilfrid Laurier University

Paul Maxim obtained his MA in criminology at the University of Ottawa and his PhD in sociology at the University of Pennsylvania where he specialized in criminology and research methods. He is currently a professor in the Department of Economics and the Balsillie School of International Affairs at Wilfrid Laurier University in Waterloo, Ontario. His primary areas of research interest are population and labour economics.

Len Garis, Adjunct Professor, University of the Fraser Valley

Fire Chief for the City of Surrey, British Columbia, Len Garis is an Adjunct Professor in the School of Criminology and Criminal Justice & Associate to the Centre for Social Research at the University of the Fraser Valley, a member of the Affiliated Research Faculty at John Jay College of Criminal Justice in New York, and a faculty member of the Institute of Canadian Urban Research Studies at Simon Fraser University. His focus is on addressing public safety challenges through evidence-based decision making and innovation.

Darryl Plecas, Professor Emeritus, University of the Fraser Valley

Darryl Plecas is Professor Emeritus in the School of Criminology and Criminal Justice at the University of the Fraser Valley. Prior to his retirement, he held the RCMP Senior University Research Chair at UFV. He received the University's Teaching Excellence Award, and in 2003 received an Innovative Excellence in Teaching, Learning and Technology Award at the Fourteenth International Conference on College Teaching and Learning. He has published numerous articles relating to Canadian criminal justice. He also has extensive experience in applied policy and program evaluation, and effective decision making.

Mona Davies, Legal Analyst

Mona Davies has an LL.B. from the University of London and her Masters in Political Science from the University of Toronto. She also has a professional degree from the Institute of Chartered Secretaries and Administrators. She has worked both in the private and public sectors with over 10 years' experience in international law, specializing in investor state arbitration.

Table of Contents

2	Forward	49	Statistics
3	Introduction	61	Experimental Designs
13	Defining the Problem	73	Program Evaluation
25	Thinking Critically	89	Costing Analysis
37	Collecting Evidence	105	Making Decisions

This project was commissioned by the City of Surrey, British Columbia, as part of its Emerging Leaders Program, which seeks to develop high performing individuals through education, workplace experience, and mentorship. An integral component of the program is the application of newly developed skills through real business projects. This edition of *The Right Decision: Evidence-based Decision Making* is an important project outcome that benefits the City, the participants, and all municipal staff who will gain from its application. Other partners include the University of the Fraser Valley, B.C.

This manual was adapted from *The Right Decision: Evidence-based Decision Making for Fire Service Professionals*, developed through the support of the Canadian Safety and Security Program, led by Defence Research and Development Canada's Centre for Security Science, in partnership with Public Safety Canada.



For their assistance, we would like to thank:

- Yalda Asadian, Special Projects Manager, Parks, Recreation and Culture
- Tammy Britton, Technical Lead, Finance & Technology Department
- Ron Gill, North Planning Manager, Planning & Development Department
- Trent Hatfield, Environmental Technologist, Engineering Department
- City of Surrey Emerging Leaders Program
- Alex Tyakoff, Strategic Planning Analyst
- Vivian Zhang, Civil Engineering undergraduate student at the University of Victoria
- Dr. Charles Jennings, Director of the Regenhard Centre for Emergency Response Studies at the John Jay College of Criminal Justice, New York

Forward

Government employees, whether they work for municipal, provincial or federal governments, are key to providing effective and efficient services to residents of our cities, provinces and country. This is true both for those on the front lines providing assistance to the general public, and for those who work behind the scenes, helping to develop sound policies and effective programs under the leadership of elected officials. Government service workers—civil servants—are a key part of sound government. These are people who are tasked with making government and their policies work: with helping to refine the structures of civil programs and ensuring that they are implemented as planned.

Government service workers, both as employees and as citizens, have a vested interest in ensuring government programs meet the needs they were designed to address. Typically, government employees are accountable to both the public and their political leaders. Those in managerial positions are particularly responsible for ensuring that priorities are met, and that programs are doing what they are supposed to do, and in a cost-effective manner. While this states the obvious, in practice this can be a difficult task to assess thoroughly.

The only firm way of thoroughly measuring our effectiveness and efficiency is to examine our programs and services by measuring the outcomes of our programs and policies.

In other words, what evidence do we have that we are doing the right things in the right ways? One framework for assessing this is evidence-based decision making. This strategy brings together a series of techniques under a basic approach that uses hard evidence (often in the form of data), to measure our success. Evidence-based decision making is a transparent tool that helps us become more effective in our decision making in developing, nurturing and maintaining government programs.

This manual provides an overview and introduction to evidence-based decision making for those who work in the broader government sector. It also includes an accompanying workbook with concrete case studies from Surrey that will help readers put in perspective the theories elaborated. By becoming familiar with the general approach and the techniques presented here, I think you will find that your decision making will become more effective. Data-driven, or evidence-based approaches, are also more effective ways of justifying what we are doing. By looking at objective indicators, civil servants, politicians and the public have a firmer basis for assessing the worth of our policies and programs and ensuring the public gets the best value for our efforts and their tax dollars.



Vincent Lalonde, M. Sc., P. Eng
City Manager, City of Surrey

Introduction

Effective Decision Making in a Changing World

While the primary function of government has remained consistent over time—to provide services to citizens—how various departments deliver those services is constantly evolving. Government programs have become more complex with time. Furthermore, the public increasingly demands that departments integrate their functions with one another to include more comprehensive services. These services often require more sophisticated resources, processes, and better or differently trained personnel.

As a result, leaders and managers continually face this question: How can we provide quality service in light of these demands while being sensitive to resource and economic restraints? Choices and tradeoffs need to be made, and consequences need to be considered. The pressure on decision makers increases when politicians, interest groups, and ultimately, the public scrutinize these decisions. The days are gone—if, indeed, they ever existed—where government leaders and taxpayers take a request for more resources at face value. Politicians, program managers and executives at all levels are increasingly forced to make choices within tight resource constraints.

The days are gone—if, indeed, they ever existed—where government leaders and taxpayers take a request for more equipment and more personnel at face value.

More than ever, government leaders need to make decisions in ways that are transparent and justifiable. Good decision making, we will argue, needs to be supported as much as possible by evidence, research, and sound information. We term this approach evidence-based decision making. We make and justify evidence-based decisions by referring to independently supported and verifiable facts. This approach helps ensure that the decisions we make are sound and defensible. Used effectively, evidence-based approaches can help you produce the results you seek.

So why is evidence-based research important? Why does this approach to problem solving matter for government service employees?

Among some key reasons are the following:

- Policies and programs not guided by sound evidence frequently cost too much, waste resources, or simply yield poor or unknown results;
- External decision-makers who approve departmental budgets may not view departmental requests as justified if they lack compelling evidence; and
- Policies and strategies that are evidence-based often produce better results, which can increase your credibility and support for the department and its mission as a whole.

This manual will help you understand how to find and use the information and research needed to make evidence-based decisions. It will also help you to put your decisions within a compelling framework to convince others of their merit.

Of course, not all decisions are or can be based on facts. Both professionally and in our personal lives, we refer to ethics, values, preferences and political choices. To believe or do otherwise would be to deny the complexities of social life. Yet, even in those circumstances, evidence-based decision making can help you link the values, principles, and ideologies that guide your department to independent evidence and supportive research.

The evidence that we will learn to use comes from a variety of sources. Some is available as administrative data

This manual will help you:

1. Find and use information and research to make evidence-based decisions.
2. Put your decisions into a compelling framework to convince others of their merit.

that governments and other formal organizations routinely collect. Some is generated during formal policy and program evaluations, and some will come from the work of private analysts and academic scientists. Other sources of information will include your own organization and, often, your own department.

Learning to Navigate the World of Evidence

This manual will help you to navigate the world of evidence without feeling intimidated by it. As we will discover, not all evidence or data is of equal value. Even good information needs to be placed in a context where we can evaluate its accuracy and meaning. In other words, this manual will help you figure out what you need to know about data generation without having to be a scientist or scholar.

Besides learning how to assess evidence, we will also discuss how to use evidence to formulate a persuasive argument. Data alone is not sufficient to inform and support your decisions. We need to frame public justifications for our policy or program decisions logically and coherently. Requests not grounded in a sound strategic or business plan will have very little chance of success. We will learn that many arguments or justifications that are put forward simply do not make sense. We will examine some major logical fallacies that we should avoid at all costs.

This manual will also explain how to conduct an environmental scan and a SWOT analysis (an assessment of an organization's Strengths, Weaknesses, external Opportunities, and Threats). You will learn why those frequently form part of the information-collection process before a new policy or program

is developed, or before strategic priorities are determined. You will learn about cost-benefit analyses and costing studies, which are critical components of strategic planning when resources are tight.

Using examples from government services across Canada, this manual will show you how to define a problem. It will help you to think critically and creatively about it, and find the evidence needed to inform your decision. Additionally, it will provide simple explanations of various forms of research so you will know how and when to use them to support your case.

Before we begin, though, it is helpful to consider more deeply the reasons for doing all of this. How and why has evidence-based decision making become so important? Why should you, or anyone else, care about the process?



Medicine and Health Care Services Have Led the Way¹

In the public sphere, we can trace the origins of evidence-based approaches back to the 1980s. Faced with significant financial challenges, the government of the United Kingdom started to emphasize the need for policies and best practices supported by compelling evidence and empirically sound

research. Decision makers had wasted too many resources, they believed, on choices that had little data to back them up. They too often decided based on personal preference, traditional practices, and ideas that had little more to support them than they were popular at the time.

As anyone who has been in their field for a while knows, the world is full of scam artists selling the latest managerial elixir or practice. Within the U.K., it was obvious to the government that investments were needed, but those investments needed to be effective and efficient and not just based on an untested ideology.²

This approach influenced many other fields but most particularly health sciences, where researchers could directly link poor practices to increased levels of harm for patients. Evidence-based medicine evolved as a way to reduce the gap between academic research and clinical practice. Ideally, this would ensure the best possible outcomes and the most appropriate care for patients. Researchers and health care professionals scrutinized policies and procedures to see how they could run medical facilities in more efficient and effective ways.³

The need to change existing ways of doing things in the world of medicine became increasingly apparent. For example, one major study suggested it took approximately 15 years to incorporate the results of research into recommended policy.

As a dramatic example, let us consider that the research basis underlying a cure for a particular form of cancer might already exist. However, the lag between that discovery and even partially implementing it in a clinical setting takes about a decade and a half. Even after that lengthy period, only about 40 per cent of practitioners are using that information.⁴

Meanwhile, people who could benefit from the results of that research continued to suffer or die because the information had not influenced medical practices in a timely way. Worse still, implementing the answer might be delayed intentionally if other groups saw greater benefit and financial profit in “managing” the disease rather than in actually curing it.

Within the field of health, the push toward evidence-based decision making continues to resonate. It is not only in the U.K. that it has become a cornerstone of public health policy development. The need for sound evidence-driven decision making has become recognized as imperative by policy advocates, researchers and other stakeholders world-wide.⁵

An evidence-based approach tries to use the best available information generated through research, experiments, observation, and other factual sources to influence the creation of the best decisions and policies possible. Sometimes, this can directly conflict with other forces, values and interests, as the previous hypothetical example illustrates.



Case Study

As we have indicated, a large and growing body of literature on evidence-based decision making exists in the medical field. Similarly, the use of evidence-based approaches is gaining substantial acceptance within criminal justice.⁶ Few formal examples of evidence-based decision making in government, however, make it into the public domain. Those in the field hear of anecdotal examples but most of the details of those situations do not make it into the public sphere.

One large and reasonably well-documented example, however, is the Province of Ontario's attempt to deal with issues in the nursing profession in the early 2000s.⁷ In 1998, the province created a Nursing Task Force to address a number of items, including "help[ing] Ontario retain and attract nurses, improve working conditions for nurses, and ensure nurses have the skills they need to provide care in an increasingly complex environment."

An overview of how evidence-driven elements were brought together by the Task Force is outlined in an overview by O'Brien-Pallas and Baumann.⁸ The Task Force was fortunate enough to draw on several large databases, including an administrative database from the Government of Ontario; one from the College of Nurses of Ontario relating to registration; and, data from Statistics Canada and the Canadian Institute of Health Information.

These data were combined with a series of interviews and solicited submissions from various stakeholders and a general review of literature on challenges faced by nurses in the field. Overall, this allowed the Task Force to examine supply-demand issues relating to nurses as well as concerns relating to the job environment.

As O'Brien-Pallas and Baumann summarized:

This process balanced the 'facts' provided by the researchers and the values and beliefs of participants in the consultation process. Finally, the development of an accountability framework with a similar mix of decision makers (senior government, managers, service professionals), knowledge purveyors (unions, associations, public representatives, nurses), and researchers (NRU researchers and colleagues) [was] charged with the responsibility for ongoing monitoring of implementation of the recommendations.⁹

Ultimately, significant changes were made to address personnel shortages, staff morale and issues relating to patient services.

Effective Decision Making: The Task of Good Leaders and Managers

Decision making is what leaders and managers are asked to do. Their decisions influence the direction of their units and affect the morale and well-being of personnel who work for them. Poorly made decisions increase conflict and diminish morale. Well-made decisions that lead to tangible, positive results can increase departmental success and improve morale.

Nevertheless, even when leaders and managers see the value in an evidence-based approach, several factors can get in the way. Some administrators feel pressured to decide quickly and with incomplete information, while others might use outdated information. Government can be a fast-paced environment; there is often a need for speed. However, this tendency should be governed by sober reflection and consideration of the latest data to inform decisions and better practices. Additionally, most people rely on personal experience, observation, or gut instinct when having to make a choice. As trained public service employees, our personal experiences and judgments are often valid, but they comprise only part of the picture. Cognitive science suggests that we typically see what we expect to see. The mind is poorly “wired” to deal effectively with inherent uncertainty and the challenges of handling complex, multifaceted issues in the field.

Poorly made decisions increase conflict and diminish morale. Well-made decisions that lead to tangible, positive results can increase departmental success and improve morale.

Using evidence-based research helps to ground our experiences and opinions in a broader context of information that is ultimately more convincing. Besides, practices evolve. The tools that supported the government professional of the early 20th century are not always adequate for the new millennium.

When developing a new strategy or policy, it is best to assess what you know, what others around you know, and what the research tells you about it. It is also prudent to commit to evaluating that new policy or plan after you have started it so you can generate your own evidence to show its effectiveness. That helps to advance the field as a whole. Your department's research can then inform other departments on what works, what does not, and why. Often we are reluctant to assess a program or practice because we might find that it does not work. That is not a problem. Both as individuals and as a society, we typically learn more from our failures than from our successes.

The Nature of Empirical Research

What are we really talking about when we use the term *evidence*? Unlike the evidence that might come out of a police investigation of a crime scene, evidence in this context has a specific meaning. It refers to the results of empirical research coming from systematic data collection grounded in formal assessments, experiments, or other research models. It is a systematic approach to answering a research question that generates information or facts that are replicable, observable, credible, verifiable, and supportable.

When assessing the research available to you, some of it will be:

- *Quantitative*, generating numbers and statistics, or
- *Qualitative*, generating subjective information that is helpful in determining preferences, values, or perspectives of those responding to the questions.

Either of those approaches can generate valid data. The key is in knowing when and where to use what kind of evidence, and to be able to find out whether it is adequate for the purposes at hand.

While many good sources of supporting evidence exist, academic research has the added benefit of being scrutinized by outsiders with no personal stake in the program.

This means that other independent scholars and researchers examined the research to see if it is credible and well designed. This does not mean to say that the work is either perfect or infallible. Nevertheless, it does increase your ability to trust in the results. Research must be peer-reviewed before most academic journals publish it. Some academic journals can be highly technical and very intimidating to those outside the field. Fortunately, many sources summarize significant academic findings or translate the results into everyday language.

Common Research Methods

In the medical field, the gold standard for research has been the randomized, controlled trial. Here researchers randomly assign individuals to receive various preventive, therapeutic or diagnostic interventions, and then follow up to see the effect of the intervention. One possible intervention might be no intervention at all. This enables researchers to compare the control group (which received no intervention) to the test groups, which received the various interventions in question. Drug testing is frequently done this way. In a later chapter we will examine different frameworks for collecting evidence and discuss why researchers hold the randomized controlled trial in such high esteem.

In the social sciences, having randomized tests involving a control group is also possible. For example, we could randomly assign security alarms to some homes as a test group and compare them with another random group without alarms (the control group). This is one way of answering the research question, “Do households with security alarms have fewer break-in incidents than households without security alarms?” A roads and transportation department might also run trials to determine the effectiveness of cameras at intersections or high-occupancy vehicle (HOV) lanes.

Researchers will set up such experiments to ‘control’ the many external factors that might skew (or distort) the results. This increases the validity of the research, so that you can have greater confidence or trust in the measurements and results. Researchers are also concerned about the reliability of their result—meaning: if we continued to do this study repeatedly, would we get the same results? Would we get the same results if we ran this test in a different community? Or, is it unique to this community only and, if so, why is that? Research needs to be both valid and reliable so you know the results are legitimate and trustworthy, and not a fluke or coincidence.

Making Better Decisions

By now, you probably can see that benefits exist in making decisions influenced by sound, credible research. Quite simply, if you have done your homework, it is likely you will have a better-informed decision. Defending your decision is also easier since the process is more transparent and is based on something other than your hunch, best guess, or personal opinion.

We should recognize, though, that evidence-based decision making is best suited for objective questions. As we noted at the outset of this chapter, other decisions are influenced primarily by our preferences, values, or beliefs, and are less likely linked to research.

However, the two merge when we want to find the most effective approach to address an issue in a way that ultimately corresponds with our values. For example, improving our quality of life by providing quality police services and crime reduction is a social value that provides the motivation to do things differently. Evidence-based research helps us to know *what* to do and *how* to do it.

Notes

1. Tranfield, D., Denyer, D. and Smart, P. (2003) "Towards a methodology for developing evidence-informed management knowledge by means of systematic review," *British Journal of Management*, 14: 207-222.
2. Tranfield, D., Denyer, D. and Smart, P. (2003).
3. Vishwanath V. Baba, Farimah HakemZadeh, (2012) "Toward a theory of evidence based decision making," *Management Decision*, 50: 832-867.
4. Antman, E.M., Lau, J., Kupelnick, B., Mosteller, F., Chalmers, T.C. (1992) "A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. Treatments for myocardial infarction." *Journal of the American Medical Association*, 268: 240-248.
5. This includes Canada. See, for example, Kiefer, L., et al., (2005) "Fostering Evidence-based Decision-making in Canada. Examining the Need for a Canadian Population and Public Health Evidence Centre and Research Network." *Canadian Journal of Public Health*, 96: I-1 to I-40.
6. The US National Criminal Justice Reference Service has an extensive data base of relevant material. See <https://www.ncjrs.gov/App/Publications/AlphaList.aspx#>
7. See Ontario Ministry of Health and Long-term Care (1999) Good Nursing, Good Health: an Investment for the 21st Century. Report of the Nursing Task Force. http://www.health.gov.on.ca/en/common/ministry/publications/reports/nurserep99/nurse_rep.aspx#appC ; JPNC Implementation Monitoring Subcommittee (2003) Good Nursing, Good Health: The Return on Our Investment. http://www.health.gov.on.ca/en/common/ministry/publications/reports/nursing_roi_04/jpnc_roi_2004.pdf .
8. O'Brien-Pallas, L., and A. Baumann (2000) "Toward evidence-based policy decisions: a case study of nursing health human resources in Ontario, Canada." *Nursing Inquiry*, 7: 248-57.
9. O'Brien-Pallas, L., and A. Baumann (2000) p. 254.



Defining the Problem

Not all Decisions are Alike

We make hundreds of decisions daily, ranging from what to have for breakfast, to deciding in which room to hold a meeting, to whether or not to buy a new car. Many of those decisions are informal and relatively insignificant, and have few consequences, while some incur a degree of risk or uncertainty. Each of us takes risks every day. For most of us, reasonable risks don't prevent us from doing our daily tasks and routines. Working through a formal process to address those issues would typically be too time and resource consuming, and cause our lives to grind to a halt.

On the other hand, we face significant decisions in our personal and our professional lives when the consequences are not small, or when we need others to be engaged in making the decision. Examining the issues in detail and working through a formal process in those instances is worth our time and effort. Generally, that formal process involves creating a clear definition of the problem, outlining the alternatives, and weighing the costs and benefits associated with selecting any of the alternatives.

Evidence-based decision making can help us in those circumstances where we need to make an economically, socially or politically significant decision.

An advantage of evidence-based decision making is that it allows us to use known results to estimate a measurable outcome. The good news is that anything can be measured. No matter how "fuzzy" the measurement is, it's still a measurement if it tells you more than you knew before.¹ One can never know the actual consequences of a decision before the event. However, by drawing on experience and the available evidence, generating a reasonable and defensible expectation of a specific outcome is possible.

All of us will make decisions that result in undesired outcomes at times. That is a reality of life. The fact that we made the wrong choice is different from making a bad decision. There is a difference between not making the correct decision and bad decision making.

As we will outline, bad decisions are avoidable. Bad outcomes from good decisions, however, are events over which we might have little control. So what then, distinguishes a good decision from a bad decision? Simply, good decisions are ones that flow from where the problem is clearly articulated. They are ones where we bring as much of the appropriate and available evidence to bear as possible. A good decision is one where you can look back and with a clear conscience assert that under the same circumstances, and with the same evidence, you would make the same choice.

While getting a less-than-ideal outcome from a good decision is unfortunate, an advantage of having made a good decision is that we can draw lessons from it.

If the decision making process is transparent, understanding why it resulted in a negative outcome is possible. Did we make some incorrect assumptions? Were we missing some important information? Was our logic flawed?

In this chapter, we will consider the following:

- What is the issue and how do we problematize it?
- How can we identify the options and alternatives?
- How can we think creatively to generate new ideas?
- How do we generate alternatives?

What is the Issue?

Typically, even rational, systematic decision makers will start by making a list of alternatives. Lists are good and they definitely have their place. Nevertheless, as John D. Rockefeller once said in a different context, “A list is not a plan.”

Before we start to generate options, we need to ask: What is the purpose of the decision? What is our intended goal? Those questions are embedded in an analysis of the problem. The framework

of that analysis is generally a strategic plan or a business plan. Making a decision without planning is common. But, as the old adage goes, “if we fail to plan, we plan to fail.” Without an explicit plan, however, we generally do not know if an undesirable outcome is the result of a bad approach or that we encountered new or different circumstances. An open and formally structured process allows us to accumulate knowledge so that we are less likely to make the same mistake in the future.

Often, unplanned decisions do not end well. Planning allows us to decide logically and systematically. Proper planning makes decision making simpler and it also makes it transparent. That is, we can show critics that the choice we made was rational and reasonable under the circumstances.

When we ask the question, “What is the issue?” we are essentially asking, “How does our decision fit into and advance the mandate of our organization?”

Before Doing Anything, Ask “Why?”

Too often, we find ourselves backed into a corner when confronted with the seemingly simple request about whether we should choose Option A over Option B. This is a popular strategic move by someone who wishes to force an issue. For example, an employee may ask for a meeting to discuss performance and salary. As an opening gambit, the employee might ask, “Are you going to give me the same raise as last year or will I also get the promotion I have coming in recognition of my service to the company?”

Clearly, the employee is attempting to force a false choice. In this instance, we call it a false dichotomy because the question assumes that only the two options A or B are possible. In fact, many options may exist—the employee could be transferred, let go or simply get nothing. Let us assume, however, that this is a standard issue of performance evaluation for a reasonably good employee.

Proper planning makes decision making defensible even when the results are not as expected, and in an environment of increased police accountability, this is crucial.

Before considering the many possibilities, assessing the employee’s contributions to the organization is a good starting point. Ideally, the organization should have a performance assessment policy in place. Lacking that, however, you might ask five Ws. *Why* should you be rewarded based on your performance? *What* have you contributed to enhancing the effectiveness of your department? *Where* can we see evidence of your contributions? *Who* in your department have you helped or supported this year? *When* can we expect to see the returns on your performance?

Perhaps these are not always the most appropriate questions to ask in the circumstances, but you get the idea. The notion is to tie the request back to the goals of the department or organization and to ensure that the choices we are considering are consistent with those goals. Typically, we are trying to ensure the basis for the choices are not irrelevant. Decisions to reward employees simply because they are friendly, consistently show up for work on time, or always dress neatly are difficult to defend.

When all else fails, ask yourself, “Can I defend my decision to others in the organization, my boss, or the public?” As a former colleague once said, “I make every major decision assuming it will appear on the front page of tomorrow’s newspaper. If I can accept that, then I have likely made a reasonable choice on reasonable grounds.”

That mandate is normally part and parcel of our strategic plan. Sometimes it is embedded in our operational plan or standard operating procedures (SOPs).

As an example, an interest group might push for additional recreational services for seniors and children. At face value, this is a worthy request, but it is only one of many worthy requests that might be put forward. The goals outlined in a city’s or municipality’s strategic plan, however, might say that ensuring transportation and basic infrastructure needs are the community’s priority. Consequently, the immediate needs of the city might be upgrading the city transit bus fleet or constructing a drainage system due to a flooding problem in the city. By referring to a planning framework, we can see that focusing on seniors’ or children’s recreational services is not a high priority. Furthermore, the incremental investment in that area might provide little, if any, improvement in the municipality’s overall quality of life compared with infrastructure investment.

By embedding the decision within the framework of a pre-existing plan—such as a strategic plan—the choices made are defensible on strategically assessed grounds.

The issue, however, is whether the requested investment fits with the defined needs of the community. The issue is not one of failing to advance the overall recreational services of the community; the issue is really how best to address the needs of the community based on fixed resources and competing demands.

Undoubtedly, the manager could have listed the many requests brought to a council and the most popular alternative among those options could be selected. The point, however, is that the recreation decision was not the only one to be considered. The key here is to refer to the organization’s operational focus or, ideally, its strategic plan.

Again, by embedding the decision within the framework of a pre-existing plan or operational framework, the choices made are defensible on strategically assessed grounds. In that case, a delay in setting up a new program to provide for recreational facilities is justifiable.

Generating Ideas

Often, choices appear obvious. Do we spend more on equipment or personnel? Is our data processing equipment at the end of its working life expectancy or not? In other instances, the alternatives are not always self-evident. It is not always an A or not-A choice. In later chapters, we will examine how to conduct environmental scans and SWOT (strengths, weaknesses, opportunities and threats) analyses. These are relatively formal procedures that systematically review what others have done or might do in similar circumstances.

Before resorting to those approaches, however, several more modest ways exist to generate alternatives. You might want to consider the following options.

Talk to people outside your normal circles

Too often we limit our social and professional circles to those we already know or with whom we work. Often, this generates a group-think mentality where we reinforce the belief in a limited number of options. Furthermore, colleagues and subordinates may be more concerned about reinforcing what you have said or telling you what they think you want to hear rather than offering unique suggestions. Outsiders, however, may face similar situations but approach the issue entirely differently.

Engage in a group brainstorming session

Possible group-think tendencies aside, sometimes the people around you are the best source of ideas. They know the organization and understand the problems. Besides, they are less expensive than consultants since they are already on payroll. Ask for individual suggestions. Sometimes a group session, where we ask people to come up with “crazy” alternatives, is effective.

The semblance of a little competition can sometimes unleash new ideas. Remember, today's innovations were yesterday's impossibilities. Brainstorming can be either informal or structured. The intent is to generate as many ideas as possible and seek solutions to vexing and persistent problems.

Read more books and journals; surf the web

The more you read, especially outside your area of policing, the more novel ideas you are likely to come across. Business books are an obvious choice but sometimes great ideas come from works of fiction. Most of us like to stretch ourselves. Professional journals are a good way of keeping up with new trends. As always, the internet is anarchy and generally fits the adage that you get what you pay for. Still, gems are to be found and modern search engines are amazingly good at ferreting them out. As Stephen Covey stated, it's always wise to “sharpen the saw.”²

Focus on the people you serve—both internally and externally

Look at the world from the perspective of the people you serve both internally and externally. How they see your organization is probably very different from how you and your immediate colleagues see it. Besides the people you serve, other great sources of ideas are your partners such as professional associations, community groups, educational institutions and other areas of government. Often these connections have something of value to offer. Understanding the outsider's view can pay huge dividends.

Hire a reputable consultant

Often, you are the local expert at your core business or activity. That is why you are in your position. On the other hand, not all of your decisions relate to your core business function. Most businesses engage outside design firms, marketing agencies, web designers or management consultants. The key is to identify the area of expertise that you require. Once done, ask your associates if they can recommend a consulting firm or individual. Usually, smaller firms are more creative and less costly, but creativity is a business like any other and services can be purchased.

Of course, you need to be willing to be open to new perspectives. Don't let your prejudices get in the way. Just because you have a low opinion of someone does not mean they have bad ideas. Also, do not feel intimidated because someone can generate better ideas than you. Especially if that person is a subordinate, you automatically get credit for being smart enough for having such a creative employee on your team.

Finally, be willing to accept that sometimes, the best options are the obvious ones. A consultant who gives you a report that tells you what you already know, may not simply be lazy or uncreative. It could be that what is obvious to you is indeed the best option. Consider that your suspicions have been confirmed.

Get a Plan

Whatever its size or complexity, every organization can benefit from having a plan. Whether we term it a strategic, organizational, or business plan, the point is the same: an organization needs to know why it is doing what it is doing, where it is going, and how it intends to get there.

Without a plan, people make decisions arbitrarily. At best, those decisions will lack consistency and, at worse, they will be contradictory. A plan does not guarantee organizational success or efficiency. Not having one, however, invariably dooms an organization to mediocrity or failure.

Much material outlining how to put together an organizational plan is available both in bookstores and on the internet. Topics range from project management practices and principles to the latest in major case management. Time spent reviewing some of that material would be a good investment.

Essentially, a plan consists of four elements:

1. A general statement of organizational values.
2. A statement of goals and objectives.
3. An outline of how the organization intends to carry out or achieve its goals.
4. An indication of how to measure success.

Four elements of a plan:



Plans vary in complexity but there are advantages to keeping it simple. Complex plans are often difficult to remember and can be highly constraining. As most battlefield generals know, once the action starts, little goes as expected. Often, the best one can hope for is that the troops know what they are fighting for, that they remember the overall goals and objectives, and that the line officers are sufficiently trained to react to unexpected tactical challenges and setbacks. Thus, there is a lot to be said for keeping things simple.

Well-crafted mission, vision or value statements can be inspiring; poorly crafted statements do little more than provide a source of levity.

Statement of Organizational Values

It is currently in vogue among management gurus to spend a great deal of time identifying the fundamental values underlying our organizations. Typically, we outline organizational values in one or more of: a mission statement, a vision statement, and a values statement.

Well-crafted statements can be inspiring, and make for eloquent poster boards that can be placed on office walls and in annual reports. Poorly crafted statements do little more than provide a source of levity. As always, the best practical advice is to keep things simple and straightforward. Simple, unambiguous statements are easy to remember and easy to follow.

Essentially, a statement of values should outline the reason for the organization's existence. This is known as the mission statement. For many organizations, such as transit services, the mission may be obvious. Its *raison d'être* is to provide efficient, affordable transit services to the community. The mission statement is where you answer the great existential question, "What is your purpose?"

Value statements should also provide some expectation of where the organization plans to be in the next three to five years. What, in other words, is the midterm vision for the organization? Perhaps you see yourself as becoming the regional standard for performance.

Finally, a values statement suggests something about your core beliefs. These are meant to be foundational and inspirational. For Google, it was, "Do No Harm." In your case, it may be, "Serve the Community." While this might seem trite, it is useful to recall the core values when decision making starts to focus too much on what is in the best interest of the organization rather than the client. In this instance, what you do is not about the organization, it is about serving your community.

Statement of Goals and Objectives

An organization's statement of goals and objectives contains the targets it sets for itself. Organizational goals are the broader targets for which one is aiming; objectives are the midterm steps one sets to achieve those goals. Broad goals may be such things as providing accessible and

affordable housing, poverty reduction and providing for a safe community.

To achieve the goal of poverty reduction, listing objectives that form a series of intermediate steps is often necessary. For example, one objective might be to develop an advocacy strategy and support for primary service providers.³



Napoleon's goals and objectives

Sometimes it is easy to confuse the concepts of goals and objectives. Too often, the two are used interchangeably. While related, the two are distinct notions. A good example is to consider Napoleon Bonaparte's intentions in 1799.

Goal	Objective
Rule all of Europe	Become head of state in France
	Conquer Italy
	Conquer Spain
	Defeat Prussian Army
	Defeat the Austro-Hungarian Army
	Incorporate Poland into the French Empire
	Conquer Russia

Ironically, Bonaparte achieved all of his objectives except for the last. Despite this impressive achievement, he ultimately failed to achieve his overarching goal. He failed to consider the impact of Russia's brutal and unforgiving winters.

Implementation Procedures

As we noted earlier, a list is not a plan. Simply outlining the organization's goals and objectives is a necessary part of, but not a complete planning process. A true plan involves a discussion of how we can carry out the goals. What is the mechanism or what are the procedures that are being put in place to meet the desired outcomes? For example, one objective may be to reduce on-the-job injuries. We may link this to the overall goal of increasing worker safety.

The important issue under consideration is: How do we make this happen? Obviously, the mechanism we choose will depend upon the circumstances. Perhaps more resources should go into training personnel. On the other hand, people may have adequate training but they may not have sufficient opportunity to practice the procedures. Another mechanism might be to enhance information-sharing and working relationships between the staff responsible for safety and line workers.

This applies to all of the goals and objectives identified in the plan, whether they are "soft" objectives, such as increasing employee morale, or "hard" objectives, such as reducing work-related injuries or damage to public property, buildings or structures.

Too often, strategic and business plans identify what the organization intends to achieve, but not the means by which it hopes to meet those intentions.

Implementation procedures are the actionable items in our plan. Too often, strategic and business plans identify what the organization intends to achieve but not the means by which it hopes to meet those intentions. Put another way, if goals and objectives are the nouns in a sentence, implementation procedures are the action components or verbs.

Measuring Outcomes

Measuring outcomes is essentially keeping a scorecard. Before you can do this, however, it is necessary to show within your plan what specific performance indicators you are going to use. You should closely link those indicators to the specific objectives you have identified and, in a general sense, to the overall goals outlined in the plan. As the eminent management guru, Peter Drucker, once stated: "What gets measured gets managed."

Obviously, clear quantitative measures are easiest to use, such as changes in calls for service, the number of applications for welfare, the number of traffic accidents, or changes in crime rates. However, we should not overlook qualitative measures. Indicators of community satisfaction or fear of crime, for example, may be hard to quantify but are crucial performance elements for any service provider.

Typically, outcome measures will cover a spectrum of issues, ranging from internal performance metrics, to levels of service provision, to financial accountability. Many discussions on strategic plans suggest creating a table where we list operational objectives in one column and their corresponding measures of success in the next. These linkages are judgment calls, but complex objectives usually require more varied indicators than simple, one-dimensional measures.

Because goals are longer term and higher level notions than objectives, identifying specific measures is often more difficult. Furthermore, goals often require a more qualitative assessment than do intermediate objectives. One thing to keep in mind, however, is that while there ought to be consistency between the outcome measures of objectives and goals, there need not be a perfect correspondence.

It is possible to meet most or all of one's objectives but not one's goals. Similarly, the failure to meet one or more objectives does not necessarily mean that the organization has missed its overall goals. Practical strategic or business plans sometimes contain other items or provide more detail on certain dimensions.

We might also put details in place about what forms the organization's "value-add" for your community, or how it differs from similar organizations or service providers. Whether these items are relevant depends on the particular environment and circumstances in which the organization finds itself. Regardless, those components become part of the crucial list of elements to which we refer when we need to make a critical decision.

Often we pose questions or decisions vaguely. A good decision maker will define and clarify the issue and relate it to the organization's plan. Having done that, one can then ask subsidiary questions such as: Does the issue warrant action? If so, when should we carry it out? Is the matter urgent, important, both or neither?

Evidence-based Decision Making

Good evidence-based decision making is tightly linked to an organization's plans. This does not mean that occasionally we will not make important decisions that are beyond what we planned to do. Environments change and new issues arise while we have to be able to maintain a focus on our core business functions. The world is not static.

Effective managers, however, are sufficiently flexible to deal with those situations. Regardless, going through a planning exercise often provides a broad enough perspective or sufficient guideposts that "out of the blue" challenges can be placed within the general framework of our plans.

The primary benefit of a good plan is that it allows decision makers to be able to justify how and why they are assessing the choices they are considering. Raising the criticism that certain options have been considered is easy. In fact, for many decisions there may be an almost infinite list of possible options. We can reduce that list substantially if we point out that the suggestions may have merit, but are outside the realm of the strategic plan.

A good plan, then, lets us know what questions or issues are relevant, what options are worthy of consideration, and consequently, what evidence we need to consider in weighing those options.

Notes

1. Douglas W. Hubbard (2014) *How to Measure Anything: Finding the Intangibles in Business*. New Jersey, NJ: John Wiley & Sons.
2. Stephen R. Covey (1989). *7 Habits of Highly Effective People*, New Jersey: Simon & Schuster.
3. City of Cornwall Strategic Plan, 2013.
<http://www.cornwall.ca/en/cao/resources/CityCornwallStrategicPlan20132015.pdf>

Thinking Critically

Clarity of Thought

Evidence and data alone are not sufficient for making good and useful decisions. How we formulate an argument or explanation is just as important as the quality of the information we might bring to bear. When we consider evidence-based decision making, we need to keep two aspects in mind. First, as in making any type of case, the underlying arguments need to be based on sound logic. An argument that can lead to more than one conclusion is generally not very useful. Second, how most people think evidence or proof shores up an argument is typically not the most powerful way of making a case.

Two things seem to characterize humanity. Those are that people like to argue and, even when someone shows another that their position is false or illogical, that person generally won't change their world view. Humans are stubborn beasts with a tendency to defend any coveted untruth against the best of reason and evidence.

Evidence seems to abound that argumentation is one of humanity's most favoured social activities. Go to any sports bar on a Saturday night and you will see what seems to be inexhaustible evidence.

Humans are stubborn beasts with a tendency to defend any coveted untruth against the best of reason and evidence.

Then, there is the internet. Its rise has provided the greatest venue for half-baked ideas, conspiracy theories and their supporters since the invention of walls and graffiti. Fundamentally, evidence and sound logic rarely sway people. When was the last time, for example, someone listened to you make a case and said, "Thank you for pointing out my logical fallacies. I see that I was wrong on this issue and I will from now on change my perspective on the matter." A positive outcome is typically one where they change the topic; a negative outcome is where they turn away muttering something about you and your kind having always been idiots.

The fact is, there are some discussions to which no solution exists, either logical or empirical. Arguments over the existence of God; who is the best looking actor or actress; or, whether Aunt Helen made the world's best muffins, will never be resolved.

Generally speaking, matters of values are issues that are based on emotional preferences. On the other hand, situations exist where evidence and rationality sway us (or, at least, some of us some of the time). Economic issues, for example, typically command our more rational sentiments. Matters of health, and life and death—immunizing your children against the measles, for instance—tend to elicit a rational response. Although, it is admitted that charlatans abound and thrive in those domains as in all others.

The focus of this chapter is on those instances where, either individually or in groups, we are willing to consider rational and evidence-based input into our decision-making processes. Since those instances appear rarely in the affairs of humans, it is obligatory for us not to miss the opportunity for making a sound decision by using faulty logic.

Logical Fallacies

Logical statements are generally of the form: if A leads to B and B leads to C, then the occurrence of A will lead to C. Logical fallacies are ones where inherent gaps, contradictions or simple irrelevancies in arguments go unacknowledged or unchallenged. Some logicians and philosophers have made careers listing almost infinite varieties of fallacies (again, see the internet). For the most part, however, logical fallacies fall into a small group. Learn to identify these and you will be less likely to be led astray, whether intentionally or not.

Appeals to Authority

None of us has the capacity to generate all human knowledge from scratch.

As youngsters we are taught that what our parents, teachers and other “experts” say is generally true. It is an accumulation of knowledge passed from one generation to the next that distinguishes humans from other beings. This has allowed us to develop antibiotics, to build skyscrapers and to distribute spam to those little boxes we call cell phones. Without accepting knowledge passed on from authorities, civilization could not exist.

However, while we may be willing to accept the received wisdom from our resident Yodas, we should not be blind to the fact that Yoda may be wrong. There is nothing untoward about asking for further evidence to back up some authority’s claim.

Be suspicious of opening lines such as: “But, it has always been done that way,” or “The experts agree that...”

While we do not have the time to question all authority, certain appeals should raise your suspicion.

Typical openings that should cause you to be suspicious are lines such as:

- “But, it has always been done that way.”
- “Everyone knows that’s the way it is.”
- “What do you (we) know? So-and-so is an expert in these matters.”
- “Science tells us that . . .”
- “The experts agree that . . .”

In such instances, there is nothing wrong with saying that, “If that is the case, then there should clearly be some hard evidence to back it up. Perhaps we should check it out in more detail.” Or, “Gee, that’s interesting because some (scientists, experts, etc.) say just the opposite. How are we to resolve this?”

Usually, appeals to authority are code for either, “I am too lazy to check this out,” or, “I am blowing smoke.”

Personal or Ad Hominem Arguments

Ad hominem is Latin for “against the person.” Essentially, ad hominem arguments are ones where someone attacks the integrity of the person making the statement. Usually, the person’s sanity, morals or parentage is called into question. An ad hominem argument is an attempt to “blow off” the proponent by undermining their credibility. Among some more polite ad hominem attacks are such statements as:

- “What do you expect from a couple of fascists (socialists, liberals, academics, whatever)?”
- “That’s a typical statement from someone who is clearly out of touch with today’s realities.”
- “That’s a typical male (feminist) response.”
- “Gee, you would think s/he is an expert in the matter the way s/he is going on.”
- “So, how many years have you been in the field?”

The key here is to separate the argument or assertion from the speaker. Just because one has a low opinion of the other person, doesn’t necessarily mean that what they have to say is wrong or irrelevant. It may be difficult at times, but trying to respect the idea is essential, if not the person presenting it.

The “Red Herring”

Red herrings are irrelevant issues that someone brings up in a discussion. For example, someone asserts in a council meeting, and it may be the case, that too much money is being spent on travel, toys for administrators or overtime. Someone then suggests that this would not have happened if we had invested in the appropriate technology a couple of years ago.

The problem here is that unchecked and inappropriate spending is due to a lack of financial oversight. Effective financial oversight has existed before the time of the Romans and long before computers were available. Investing in the appropriate technology may help in the oversight process but does not ensure oversight in itself. Examples are bountiful of solutions that have merely added to the problem rather than solving it.

The key to addressing red herrings is to ask how the herring is related to the problem being considered. How will the technology be used to enhance oversight? Is the appropriate software available? Are the auditors properly trained in the equipment to be able to enhance their performance? Computers, after all, only do what we tell them to do.

Pink Herrings

True red herrings are items that are clearly unrelated to the issue at hand. Sometimes, however, someone may raise an issue that is sufficient to address the problem but is not necessarily a solution. We might refer those to as pink herrings. Perhaps the biggest pink herring is for administrators to argue that the problems exist in their organization because of a lack of financial resources.

Money can purchase resources. All too often, however, more money just leads to more of the same. Money, itself, doesn't necessarily solve the problem. Proper oversight, a more effective use of existing physical and human resources, or a more creative approach to the issue may be more effective than simply throwing more money at the problem. What is necessary is that existing or future resources are directed toward developing or enhancing mechanisms related to the problem.

As with the red herring, we need to address the open-ended call for money by asking how the money will be used. The answer will likely be to purchase more equipment or hire more staff. The subsidiary question then becomes: In what way will that equipment or the staff enhance a process that is currently broken or ineffective?

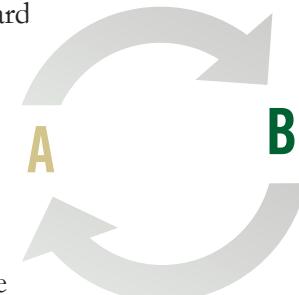
Circular Arguments

Circular arguments are those of the form that A causes B because B is the result of A. Circular arguments abound, particularly in political debates. A favourite of teachers is students who come after an exam and assert that they can't get a C because they are A students. (So, explain how you earned the C if you are an A student?)

Another good example is sometimes found in salary negotiations. Bargaining units will sometimes insist that they need to get a larger increase than their colleagues because they have historically been the highest paid unit in the group of comparable organizations. If you don't give the raise, how can they be the highest paid? Usually, most ratcheting effects that we see in labour negotiations are based on circular reasoning. Group A has it in their contract that they are to have a 10 per cent premium on the rest of the jurisdiction because of the high cost of living in their area. Group B argues that to remain competitive, they need to be within 10 per cent of Group A regardless of productivity or other factors. A change in the compensation of any one group automatically ratchets the pay of the other.

Sometimes we use the term *begging the question* to describe a circular argument. The form of the argument is essentially the same: "You know, the reason that action is illegal is because it is against the law." Being "against the law" is a synonym for "illegal," so one is simply asserting that something is illegal because it is illegal.

Similarly, an often heard comment in city councils is that a particular group will not support tax increases because they have made it part of their platform. When asked why that is part of the platform, the answer is that tax increases are not supported by the people.



To break the circularity, we need to know why a body passed the law in the first instance: what was its supposed purpose? Likewise, we need to know in what way not increasing taxes benefits the electorate. It may be that not increasing taxes denies a much needed service which is clearly in the interest of the taxpayer. In cases like this, we need to ask, what is the exact economic mechanism supposed to be at play?

Other Fallacies

People call upon many other logical fallacies when rationality and evidence fail them. They range from the teenager's perennial appeal to popularity: "But everyone at school has one," to appeals to nature: "That is just not natural." Parallels, of course, abound in the professional sphere. Every municipality or department in the region has a Nouveau Widget so, obviously, we need one too. A current bureaucratic favourite is the rationale for why we keep a current practice or why things don't change. The cliché du jour is: "It is what it is," which has replaced the formerly abused, "Well, that is the nature of organizations." All of these are logically non-starters that get us nowhere.

Just because two things appear associated, doesn't necessarily mean that one causes the other—or that, in fact, they are causally connected in any way.

Causal Linkages

One clinker of a fallacy we did not discuss previously goes by the formal name of post hoc, ergo propter hoc, which means “after this, therefore because of this.” Those who might have studied statistics will recognize this as a variant of the “correlation does not prove causation” fallacy.

Just because two things appear associated, doesn't necessarily mean one causes the other—or that, in fact, they are causally connected in any way. The possible absurdity of assuming that because two things are correlated because they are connected is presented in the fire engine fallacy. The story here is that a Martian comes to Earth and notices that wherever a fire occurs, there is invariably a fire engine at the scene. The Martian, therefore, erroneously concludes that fire engines cause fires.

Of course, this fallacy can also be applied to police cars and crimes as well as ambulances and injuries.

Obviously, association or correlation is somehow related to causation. The question is how can we identify or recognize a causal relationship when we see one? The issue is important because causal thinking and causal imagery have become entrenched in our everyday view of the world. Whenever we see something we do not quite understand, our first inclination is to ask, how did that come about? In other words, what was the cause?

From an historical perspective, formal causal thinking is a relatively recent idea. Most scholars use David Hume's writings as the starting point for explaining what is a cause and how we might identify one.



David Hume (26 April 1711 – 25 August 1776) was a Scottish philosopher, historian, economist, and essayist known especially for his philosophical empiricism and skepticism. He was one of the most important figures in the Scottish Enlightenment, and in the history of Western philosophy. He is the philosopher “widely regarded as the greatest who has ever written in the English language.” Hume is often grouped with John Locke, George Berkeley, and a handful of others as a British Empiricist.

Hume was a Scottish philosopher who lived in the early to mid-1700s. Without belabouring the issue, Hume identified three necessary conditions for a causal relationship. The first condition is that the cause and the effect must be coincidental or “conjoined,” as he said. This is the correlation part, where two things generally appear together.

The second condition is that the cause must come before the effect. Therefore, if the Martian had been around a little longer, he would have noticed that the fire occurred first and that the fire engine generally turned up later. Thus, it was the fire that caused the firefighters to respond; the fire was not a consequence of the existence of fire engines.

The third element of causation is the most difficult issue and that is what we call the condition of non-spuriousness. Non-spuriousness means the cause is not just enough or sufficient to cause the effect, but that it necessarily produces the effect or outcome. This is sometimes easier to understand in the negative. What non-spuriousness means is that no third factor is resulting in the apparent cause and effect to be appearing together. An example here might be the strong correlation between the amount of crime, the number of police officers and the population across jurisdictions. Neither the number of officers nor the number of crimes in a jurisdiction may be a cause of the other; both, however, are likely driven by an underlying increases or decreases in population density.

Hume's conditions for a causal relationship

1. The cause and effect must be coincidental.
2. The cause must come before the effect.
3. There is no underlying third factor resulting in the cause and effect to be appearing together.

Spuriousness means that a relationship between two or more factors is coincidental. The real cause is an underlying third factor. The problem here is that even if we take away the apparent cause, the effect will remain. Thus, with crimes and police cars, if a prank caller instigates a call that makes police cars appear, then they will appear whether a crime occurs or not. From an evaluator's or a scientist's perspective, non-spuriousness is generally the most difficult factor to control. Observing that two events generally coincide is not difficult, nor is it difficult to see that one event generally precedes the other.

The difficult issue is assessing whether some other underlying mechanism is driving both of those events. Essentially, we have devised two ways to deal with the spuriousness issue. The first is to try to develop explanatory theories to explain how or why something should cause something else. In formal terms, we need to find what we call a causal mechanism. Logically, why should X produce Y? As we say in the trade, “What’s the story?”

The second way of dealing with the non-spuriousness issue is through the physical manipulation of conditions. That is, can we physically reproduce the effect ourselves? We call this manipulation an experiment.

Over time, we have developed a series of experimental designs or ways of manipulating situations so that we can isolate what we believe are the cause and effect factors from other possible or spurious influences. We will highlight those techniques in a later chapter.

In summary, then, it is sufficient at this point to consider that all three conditions must exist for us to be reasonably confident that something is truly the cause of something else. Those are the elements of coincidence or correlation; temporal sequencing where the cause precedes or comes before the effect; and, the condition of non-spuriousness where no other underlying mechanism is generating both the apparent cause and the effect.

Unfortunately, we conduct much research that does not consider all three of those issues. That is why, for example, we often hear of some medical survey where some factor (say, pomegranates) is supposed to reduce the risk of cancer. Typically, the study is correlational such that someone conducts a survey and it is found that people who eat pomegranates have a lower incidence of cancer. We can probably determine that the consumption of pomegranates preceded the onset or non-onset of cancer.

What those studies generally do not do is to control for spurious or confounding factors. For example, pomegranate eaters may be also less likely to smoke, get more exercise, eat a healthier diet and generally have a healthier lifestyle than non-pomegranate eaters. Those factors are likely the real causal agents. Including pomegranates in the diet or not is irrelevant.

Of course, once we start to believe that pomegranates are related to cancer, we can generate any number of possible causal explanations after the fact. For example, we might argue that high levels of vitamin C or antioxidants in pomegranates fight the onset of cancer.

Linking Evidence to Explanations

A common mistake people make is to think that by collecting sufficient evidence, one can “prove” that a hypothesis or theory is correct. In fact, the relationship between an explanation and what forms evidence is complex.

To prove a relationship, we generally need to use data or evidence in two ways. First, when we consider an explanation, we must find one that is consistent with at least most of the evidence or facts that we have to date. If an explanation does not explain most of what we know, it is unlikely to be a good candidate for what we need.

Once we have narrowed our plausible explanations to ones that make sense logically, and ones that generally fit the existing evidence, we need to conduct secondary tests to see whether those explanations hold up under critical circumstances. Obviously, we have selected an explanation that fits the known facts, so simply collecting more data under the same circumstances likely won’t give us more hard evidence.

For example, the fact that crime rates in inner-city neighbourhoods with graffiti tend to be higher than other neighbourhoods does not provide proof that graffiti causes crime. Going back to our Martian example, seeing ever more instances of fires and fire engines appearing together does not provide more

To prove a theory:

1. We must find an explanation that is consistent with at least most of the evidence we have to date.
2. We must then conduct secondary tests to see whether those explanations hold up.

proof that one causes the other. On the other hand, a few instances where fire occurred with no fire engine nearby soon disproves the hypothesis.

That is perhaps the single most important point that Hume made in his discussion of causation. It is very difficult to prove something is true; it is much easier to show that it is not true.

One example Hume used was that just because the sun has risen in the east and set in the west since time immemorial, it does not “prove” that this will necessarily happen tomorrow. On the other hand, all we need is one instance where the sun doesn’t rise in the east to disprove the pattern. As contrived as that example might be, it does make the point about the relative imbalance between evidence that appears to show a relationship, and evidence that appears to dispel a relationship.

Working and Null Hypotheses

Hypothesis testing is an approach routinely applied in science to help establish knowledge. It requires a true (or false) statement to be made that offers a plausible explanation about the problem. Testing the hypothesis results in our coming to some conclusions.¹ For example, a working hypothesis might be that initiating congestion fees in the downtown core during peak hours will deter people from driving their personal vehicles and ease traffic congestion. This seems to make intuitive sense and would appear consistent with basic economic theory. On the other hand, it might be that congestion fees have no impact. Support for this side of the argument comes from the notion that many people may be willing to pay the fees for the convenience of driving their own vehicles. Furthermore, some drivers, such as delivery truck drivers, have no choice in the matter and will make deliveries and clog traffic regardless.

To provide evidence of whether our working hypothesis is really so, we would test the hypothesis by looking at instances where the opposite could be the case. That is, where fees have had no impact.

This leads us to what we call the null hypothesis: that is, there is no statistically significant difference among instances where congestion fees are implemented and where they are not. If we fail to reject or falsify the null hypothesis then we must logically reject the working hypothesis that congestion fees really do work.

It is this strategy that scientists use to test hypotheses and theories. We cannot prove the working hypothesis directly. Instead, we create a null hypothesis that is the opposite of the working hypothesis. If we find support for the null hypothesis (that is, we find that congestion fees have no influence whatsoever on the outcome) we toss out the working hypothesis. Or, at least, we need to seriously reconsider what it says. Perhaps in this instance, the fees are simply not high enough. If we do not find support for the null hypothesis of no impact, we have very strong reasons to believe that our working hypothesis is valid. As we find that fewer and fewer alternatives pan out, the greater credibility we have in the working hypothesis.

Case Study: Peer-based Needle Exchange Programs

One of the significant challenges local health boards face is the spread of HIV/AIDS, hepatitis and other infectious diseases among intravenous drug users. Among most professionals, the feeling is that the secondary impact of injecting drugs often has significantly more adverse effects on the drug user than the drugs themselves.

In order to minimize or mitigate the impact of injecting needles, many jurisdictions have created needle exchanges and so-called “safe” injection sites. While the literature on syringe exchange programs suggest that they have a significant impact on reducing HIV and other blood-borne diseases, it appears that there are significant subgroups of users who do not participate in these programs.

In Vancouver, B.C., a peer-run outreach network was set up by the Vancouver Area Network of Drug Users to complement the more traditional exchange program model. Kanna Hayashi and colleagues conducted an assessment of whether this peer outreach program had the desired or expected impact.²

Working hypothesis:

In this particular instance, the working hypothesis was that the peer-run program would reduce the reuse of needles among intravenous drug users who typically did not access traditional exchange sites.

Null hypothesis:

The null hypothesis would be that the use of a peer network would have no impact on needle reuse.

Both logical and behavioural explanations to support both the working and the null hypotheses could be put forward. Support for the effectiveness of peer intervention largely rests on the notion that the peers were mobile, understood the behaviour of the hard-to-reach group, and could deliver clean syringes where they were needed. Support for the null hypothesis is based on the insight that these hard-to-reach groups were difficult to locate (typically not having stable housing) and used drugs that would make them exceedingly reluctant to understand the value of clean needles. Ultimately, however, Hayashi et al. determined that “access to this service was associated with lower levels of needle reuse.”

Notes

1. Chainey, S. "Improving the Explanatory Content of Analysis Products using Hypothesis Testing", *Policing Advance Access*. March 14, 2012.
<http://policing.oxfordjournals.org/content/early/2012/03/14/police.pas007.full.pdf>
 2. Hyashi, K., E. Wood, L. Wiebe, J. Qi and T. Kerr (2010) "An external evaluation of a peer-run outreach-based syringe exchange in Vancouver, Canada." *International Journal of Drug Policy*, 21: 418-421.



Collecting Evidence

Environmental Scans

The plans we create help us set priorities, influence evidence-based decision making and affect our organization's ability to fulfill its mandate. When issues arise and decisions have to be made, we need evidence to help us decide the likely impact or effectiveness of our decisions. Government departments can use this approach to improve their performance and stay ahead of public expectations.

A common strategy for gathering this information is through an environmental scan. Simply put, an environmental scan gives us an informed, comprehensive picture of the current circumstances in which our organization operates. It makes us aware of internal and external realities, important issues, and trends affecting the organization. Information of this kind helps confirm or refute our perceptions. It can guide us with future programming, strategic priorities, and budgeting. An environmental scan can also be useful in determining future strategies and in developing appropriate, well-informed responses.

What benefits do organizations receive from conducting an environmental scan? Why should we spend the time and energy to conduct one?

An environmental scan makes us aware of internal and external realities, important issues, and trends that affect our organization.

Among the most prominent are the following. Environmental scans can provide:

- A fresh, objective look at issues within the organization's goals and mandate, with an eye toward how to rank them most effectively;
- An opportunity to access information, research, statistics, and other data that someone else took the time to collect;
- An opportunity to involve community stakeholders, organizations, individuals, and groups in decisions that affect them, by giving them an opportunity to provide input, perspective, and advice;
- An opportunity to discover the strengths and assets in the larger community to address the issue;
- A framework or point of comparison to understand the assets and strengths of own organization; and

- An opportunity to learn how your organization's programs and practices are affecting other organizations, agencies, individuals, or groups, and to what degree your programs and practices are effective in fulfilling your organization's mandate.

Conducting an environmental scan is a sequential process that involves gathering information from secondary sources, including existing research reports, statistics, or other information. First-hand or primary sources of information from individuals or groups that you will contact yourself will supplement this. Analysis of this information leads to establishing where your organization fits within the broader social ecology.

Unlike many other management procedures, few formal guidelines exist for conducting environmental scans. What we will do, however, is give you an overview of the procedure and some suggested tools for moving forward.

Types of Environmental Scans

There are essentially two types of environmental scans. The first approach is a less formal type of scanning that you conduct yourself, based on your own knowledge and what you or an assistant can gather sitting at your desk. The first step is to write out what you know about how others are dealing with similar situations. In other words, you are looking to see how others in your social environment do things.

Generally, people who are more connected with their colleagues, who read the trade literature, and who regularly attend conventions and workshops often find this process easier.

A second part might involve a more formal review. Depending on the issue, you might seek out journal or news articles written on the topic. A good place to start is to check the internet. Search engines such as Google, Bing and Webcrawler can retrieve a tremendous amount of information very quickly. A big challenge in using general search engines is that identifying the exact search terms you need is sometimes difficult. Consequently, the search generates more chaff than wheat.

Using Internet Search Engines

There are some tricks to using search engines. If you are fortunate enough to have access to a municipal librarian or a local college or university library, there are usually experienced people who can provide assistance. We provide some tips for Google searches on the next page.

Either online or by visiting a library, it is also possible to search the professional literature. Trade magazines and journals often provide coverage of general issues.

Effective Searches on Google

1. Be specific.

Find pages within sites using *site:[website URL]* and your search phrase, find authors using *author:[name]*, and type *intitle:[word]* to find a page with that word in the title.

2. Format.

Use *filetype:[pdf or other extension]* to find images and all sorts of files (such as docs and jpgs).

3. Broaden your search.

Use an asterisk (*) as a wildcard search operator to fill in the blanks. For example, “*transport**” will return information on transportation, transporting and so on.

4. Limit your search by excluding unwanted terms.

Put a minus sign in front of terms you wish to exclude. For example, *alarms -burglar* will exclude the term “burglar” from your search. To limit a search numerically, use the range (two dot) indicator. For example “*used snow plows 2010 .. 2014*” will limit results to those years.

5. Use specific search engines.

Google scholar, for example, is an excellent way to find both academic and other articles on selected topics. Webcrawler looks across a series of search engines. Also check out the website for Amazines (www.amazines.com) for a database of free articles.

Speaking with a librarian well-versed in your organization’s mandate at a local university or college, or knowledgeable staff at an in-house library, is a good place to start.

For more detailed sources of information, you may need to enter the formal research or academic literature. This latter step can be a little daunting at times since there is much variation in how people write technical articles. Some articles are very

accessible while others require extensive prior knowledge of the topic. The key is not to become discouraged.

Sometimes it is worthwhile looking further afield. Here, official websites such as that of the Federation of Canadian Municipalities can offer a wealth of information. Many local universities and college programs have partnerships with government and are a great source of information and potential joint projects.

Sources of Information

Professional and trade journals are an excellent source of information. While there are few general government services journals, many areas within government, such as transportation, fire services, human resources and so on, have specialized publications. Although it is focused on the U.S., the Govloop website is an excellent starting point for a range of issues-based publications relating to government.¹

If you require information on characteristics of your community or other statistics, a great deal of information is available on the Statistics Canada website. Statistics Canada collects a significant amount of information on a wide range of subjects.² As well, some provinces collect their own data although it is often in conjunction with StatsCan. Some key provincial links are provided in the box on the next page.

If you want to look further afield, for comparison data for example, the U.K. Department for Communities and Local Government collects and publishes official statistics relating to deprivation, fire and rescue services, housing and homelessness, local government finance, planning performance and land use.³

To do a scan most effectively, make sure you have collected information in more than one way. By doing this you can check and cross-reference to see if the same issues and concerns are surfacing through your various sources of information.

Occasionally, conducting a formal process where others in the organization are involved is worthwhile. In this instance, you might consider bringing in an outside facilitator and conducting a formal scan. The process of doing a formal scan is outlined in the second part of the chapter on SWOT analyses.

The primary difference between an environmental scan and a SWOT analysis is that the focus or range of issues considered by an environmental scan is generally much broader. SWOT analyses are typically limited to issues relating to challenges and opportunities confronting the organization.

Links to Provincial Data Sources

B.C. <http://www.bccstats.gov.bc.ca/Home.aspx>
Alberta <http://finance.alberta.ca/aboutalberta/osi/>
Saskatchewan <http://www.stats.gov.sk.ca/>
Manitoba <http://www.gov.mb.ca/mbs/>
Ontario <http://www.ontario.ca/government/ontario-open-data>
Quebec http://www.stat.gouv.qc.ca/statistiques/index_an.html
New Brunswick <http://www.snb.ca/e/0001e.asp>
Nova Scotia <http://novascotia.ca/sns/access/vitalstats.asp>
Newfoundland and Labrador <http://www.stats.gov.nl.ca/>
NWT <http://www.statsnwt.ca/>
Yukon <http://www.eco.gov.yk.ca/stats/>
Nunavut <http://www.stats.gov.nu.ca/en/home.aspx>

Framing Your Environmental Scan

The question you are trying to answer influences the information that you will be gathering. To frame the environmental scan, we can start by asking some focused questions, such as the following:

- What is the key issue?
- What do we need to know about the issue?
- What are the trends and drivers affecting these factors?

Once you have framed the question, and you have gathered the research from primary and secondary sources, then the analysis begins.

First, we need to consider what themes, concepts, issues, or concerns surfaced in the secondary research. In other words, how have other groups, organizations, communities or governments elsewhere been affected by this issue? How have they ranked those concerns?

Compare the results of your surveys with the qualitative data that is emerging from your focus groups. Consider what people have been saying in the one-to-one interviews. What common themes are emerging? How are the results showing consistency and repetition?

Try to figure out how these people have ranked the concerns that also showed up in your secondary research. Do they see it the same way? Or have they raised different thoughts, ideas, or concerns that have not shown up in the secondary research?

Once you or your team have agreed on the ranking of the issues, beginning with the most serious and urgent, then you can begin to consider the strategies, program activities, and practice that will help you address them. You will also need to consider the budget implications involved in meeting these strategic priorities.

As we noted, a SWOT analysis often accompanies environmental scans, which determines the internal and external strengths, weaknesses, opportunities and threats that are affecting the organization's ability to fulfill its mandate. The SWOT analysis is explained more fully in the second part of this chapter.

An example of an ongoing environmental scan is presented on the next page.

Example: Employment and Social Development Canada: Ontario Region Labour Market, Spring 2015

Once a year Statistics Canada does an environmental scan of the Ontario labour market. The scan provides a general overview of the demographic, economic and labour market conditions and provides information to help identify potential pressures on Service Canada's program delivery and services.

The scan identified the following key points:⁴

General Overview/Economic Context

- Global economic growth was uneven and uncertain throughout 2014 making pre-recession growth seem now unsustainable.
- Clear strength in the United States (U.S.) and United Kingdom (U.K.) economies contrast mixed fortunes for Asian economies, slowing emerging markets, and a stagnant Eurozone dealing with increasing economic and political risks.
- The U.S.—Ontario's largest external trading partner—is expected to have grown by 2.4 per cent in 2014, and is projected to strengthen further to 3.6 per cent in 2015, leading advanced economies globally.
- The Canadian economy grew by an estimated 2.4 per cent in 2014 based mainly on the strength of exports, but is expected to increase by only 2.3 per cent in 2015 as impacts from the oil price decline partially offset strong exports.
- Lower oil prices could lead manufacturing-centric provinces like Ontario to overtake the western provinces in 2015 Gross Domestic Product (GDP) growth.
- Ontario's economy is expected to continue to improve with growth rates of 2.3 per cent in 2014 and 2.7 per cent in 2015 as exports, particularly manufacturing-related, receive a boost.

Continued on next page

Ontario Region Labour Market (*cont.*)

Provincial and Local Labour Market Conditions

- The Ontario unemployment rate continued to fall post-recession to 7.3 per cent in 2014, but it was mostly attributed to fewer people looking for work.
- In 2015, employment is projected to grow by about 1.0 per cent and the unemployment rate is expected to average 6.8 per cent.
- The labour market outcome of Ontario youth aged 15 to 29 improved in 2014.

- Employment growth in the services-producing industries generally fared better than in the goods-producing industries, particularly the trade (+28,000), and professional, scientific, and technical services (+17,200) industries.
- Employment growth was fastest in the Muskoka—Kawarthas (+10.6 per cent) economic region, while the Toronto region remained flat (0.0 per cent).

Based on this and other information, you might decide to refocus the service components of your department or organization. Clearly, several options are available. Depending on your location (a major metropolitan area as opposed to a smaller community), you might wish to expand or restructure your range of services.

Another option would be to identify the four or five key areas in which all other departments engage, and focus on those as your core functions. Again, what you get out of an environmental scan is determined by the initial question you are trying to resolve.

SWOT Analyses

A SWOT analysis is an assessment of an organization's Strengths, Weaknesses, Opportunities, and Threats. Keep in mind that, typically, the strengths and weaknesses are internal to the organization, while the opportunities and threats are characteristics of the external environment.

SWOT is easy to use. It can be a useful complement to the environmental scan. A SWOT can generate crucial information with relatively little effort, and it brings that information together in a framework that provides a good base for further analysis. It is an excellent decision-support tool, and aids us in making an important decision—especially the right decision.

As we discussed earlier in the chapter, the environmental scan will give you primary and secondary information to identify pressing issues and concerns related to your research questions. When that information is combined with the results of the SWOT, you will be better equipped to identify your strategic priorities and future directions.

The SWOT adds to the results of the environmental scan by engaging various members of your organization in a discussion of the strengths and weaknesses that exist within your department. Looking outside the department allows you to consider opportunities that you could seize to advance the interests of the organization.

The SWOT also explores threats: those external factors, realities, or trends that can make the ongoing functioning of the department more challenging.

We sometimes conduct a SWOT analysis as a group session with a facilitator. A survey that each member of the group completes in advance might precede this, so they have a chance to consider their own assessment before group discussion begins. Even simpler, one can give each group member a blank SWOT template that they can use to jot down their thoughts in advance, and then have them bring it to the meeting.

Conducting a SWOT Analysis

While conducting a SWOT analysis by yourself is possible, we usually see the real benefit of the exercise when several members of the organization are involved.

One of the paradoxes managers face is that on the one hand, employees and others expect leaders to lead but, at the same time, they expect to be part of the decision-making process. As with any activity, consultation has a price. While you are consulting employees, they are not doing their normal activities.

Furthermore, group dynamics can generate unexpected results. Group politics come into play and red herrings can occupy a significant amount of time. For those reasons, having an outside facilitator lead the exercise is often beneficial. The advantages of consultation, however, are numerous.

First, groups often generate crucial ideas that a single manager or even a management group might overlook. Second, people from different segments of the organization interact with different audiences, suppliers, community groups, clients or customers, regulators, and other service providers.

This gives them different perspectives on the organization, particularly with regard to outside influences. Third, even participants who do not see their input reflected in the final product generally feel they have had some say in the process. This typically has a positive effect on morale and often creates more “buy in” when choices have to be made and different options are implemented.

In a group situation, one of the first questions when conducting an analysis is: Who will participate? It is helpful to have a diverse cross-section of individuals to ensure the most comprehensive assessment.

While no guarantee, this helps to increase the likelihood that no crucial aspect is overlooked. As a general rule, the SWOT analysis should be done by no less than mid-level management, and preferably even a higher level of leadership. In addition, the analysis should include representative employees from throughout the organization. Front-line supervisors should be included. Again, while not always the case, leaders in the organization often have greater insight into those external and internal issues that need to be considered. This comes from their experience as well as their relationships with a wide variety of people inside and outside the organization.

Before starting the analysis, and filling in the matrix, it is often worthwhile providing the team with the environmental scan results to read in advance of the SWOT analysis meeting. Ensure you include the guiding research question that is behind the environmental scan and SWOT process, as that will create the framework for the discussion. Create helpful ground rules for the discussion.

Example: Town of Norwich SWOT Analysis

Below is a SWOT Analysis conducted by the Town of Norwich⁵ to examine the opinions and perspectives of local business owners and key community stakeholders.

The goal of this report is to provide a better understanding of Norwich's economic development status and potential from the viewpoint of community stakeholders.

Workforce and Education	
Strengths	Opportunities
<ul style="list-style-type: none"> • Township has a skilled local labour force for the manufacturing and commercial sectors – many local businesses are able to find qualified employees eventually • Auto industry in Woodstock and Ingersoll is providing technical trade employment opportunities for local youth • Many local businesses have intern or apprenticeship programs • Township features local schools in Norwich, Otterville, and Burgessville • NDHS offers a creative curriculum. • NDHS has an excellent student co-op program – many local businesses support the program • Several colleges and universities are in close proximity to the Township 	<ul style="list-style-type: none"> • Growth of the auto sector in the area will provide an additional source of jobs for local residents and attract a larger workforce to the area – promote the benefits of the small town lifestyle that's available in the Township • Strong community interest to maintain NDHS with recognition that a new model is needed. Possible options identified include: <ul style="list-style-type: none"> ◦ Expand the grades based on a grade 7-12 model ◦ Develop an agriculture training program at NDHS as another unique element of the curriculum ◦ Consider applying school surcharge to local residents for maintaining the school • Expand promotion of the high school co-op program to local businesses • Career days are viewed as an important way to assist youth in finding employment locally • Provide training opportunities for local businesses with respect to workforce planning and employee attraction / relations / training
Weaknesses	Threats
<ul style="list-style-type: none"> • Difficult to recruit/hire/keep some skilled professions (e.g. managers, engineers, welders) • Uncertainty of the future of NDHS and local elementary schools • Loss of school co-op program could result in reduced youth exposure to local employment opportunities and a less skilled local workforce • Limited number of jobs for youth in the local communities • Level of interest in the NDHS co-op program currently exceeds the supply of jobs 	<ul style="list-style-type: none"> • Reduction in the agriculture workforce • Out migration of skilled workers due to perception that larger companies offer more career advantages • Funding programs for student labour is important for many businesses – loss of funding could impact the availability of these positions • Loss of NDHS / school restructuring: <ul style="list-style-type: none"> ◦ Reduction of local jobs in the education sector ◦ Loss of affordable labour for local businesses that rely on students ◦ Loss of co-op program and reduction in youth exposure to local job opportunities ◦ Greater disincentive for families and businesses to stay here / move to the area

SWOT Discussion Ground Rules

- Focus on one quadrant at a time.
- Listen to understand, and acknowledge what you are hearing others say. Avoid interrupting or criticizing the contributions of others.
- Establish reasonable time limits to keep the discussion moving forward. Respect each other—it's acceptable to have differing points of view and perspectives
- Agree on how distractions such as cell phones and interruptions from support staff will be managed. We suggest that cell phones be turned off and administrative staff interrupt only for emergencies.
- Confidentiality: What can be shared outside the room? Where will the information go in the end? How will anonymity be protected?
- All team members should participate.

As the group considers the issues and concerns resulting from the environmental scan, ask them to consider each quadrant in turn to assess how they could more fully address those issues and concerns.

As you go through your SWOT analysis, keep these factors in mind:

- SWOT analysis is a subjective process, not a science. However, the quantitative and qualitative data that emerged from the environmental scan will help the participants trust that the results are well-founded.
- Keep it simple by focusing on a few issues only. If other matters emerge, you can address them later through a subsequent process. Without these limitations, the process may bog down with too much data and information.
- Be realistic about the strengths and weaknesses of the organization. Create safety and transparency so participants will be honest.

In summary, the SWOT analysis combines with the environmental scan to create strategic plans that are realistic, researched, and supported by internal personnel and external stakeholders. Evidence-based decision-making benefits from using tools such as these, leading to plans and decisions that will be solidly grounded in facts and research, and guided by a wide array of perspectives and input.

Notes

1. <https://data.govloop.com/Government/Government-Trade-Magazines/ztqm-a4sy>
2. <http://www.statcan.gc.ca>
3. <https://www.gov.uk/government/organisations/department-for-communities-and-local-government/about/statistics>
4. See <http://www.esdc.gc.ca/eng/jobs/lmi/publications/e-scan/on/on-escan-spring2015.pdf> for the full document in pdf format.
5. Township of Norwich SWOT Analysis Report June 2008. Prepared for: The Corporation of the Township of Norwich by Harry Cummings & Associates Inc. <http://hcaconsulting.ca/pdfs/2008%20Norwich%20Township%20SWOT%20Analysis%20Report.pdf>

Statistics

A Tool for Decision Making

Statistics is probably one of the most misunderstood of disciplines. Most university students dread having to study it, and most professors who teach it often do so with great reluctance. Furthermore, the topic is often reviled as a tool of charlatans. As Mark Twain once claimed, “There are lies, damned lies and statistics.” Yet, used appropriately, statistics can be one of the most useful and powerful tools in the decision maker’s toolbox.

Our suspicion is that statistics’ bad name stems from two sources. First, many people see it as an outcropping of math—with which most of us had a less than excellent experience in high school. Second, most people who teach statistics are not themselves statisticians and, while they may come to master the technical details, they rarely grasp the underlying logic. Statistics does entail some math, but most of that math is no more complicated than being able to balance one’s chequebook. The key to understanding statistics is to see it as a way of organizing and making sense of a world dominated by uncertainty. In fact, one definition of statistics is that it is the science of decision making under conditions of uncertainty.

The key to understanding statistics is to see it as a way of organizing and making sense of a world dominated by uncertainty.

What is key for most decision makers is not to get tangled in the details of statistical analysis, but, instead, to understand the fundamental principles or logic behind the activity. Those fundamental principles are few and, generally, quite simple. Once understood, however, the principles of statistics can be used to great advantage, even if one doesn’t have a detailed knowledge of the underlying math or technical aspects.

Statistics consists of two basic activities. The first is the collection of data in an attempt to describe something. The second is the use of data to help make decisions or inferences. The first activity we call descriptive statistics; the second, we call inferential statistics.

A Discussion of Measurement

We often refer to the process of observing and recording data as measurement. What distinguishes the way statisticians view measurement from most other people is that statisticians assume all measurement contains an element of error. In other words, in the world of statistics, having something measured with 100 per cent accuracy is more good luck than good management. From a statistical perspective, error in measurement has two basic sources: inherent error or instability, and operational error.

When we speak of inherent error or instability, we are referring to the property of the thing we are measuring. For example, if you were to ask someone to tell you on a 100-point scale how satisfied they were with their job (assuming 0 is total dissatisfaction and 100 represents total satisfaction), they might respond 71. If you asked the person the same question on several different occasions, they would likely give you a range of answers somewhere close to 71.

The reality is, most people have a general idea of their level of job satisfaction but have a hard time giving a precise number. Furthermore, while they may be mostly satisfied with their job, their exact level of satisfaction would vary according to numerous factors, ranging from the time of day, to whether they just had an altercation with their superior, to the weather.

Inherent error relates to what we are measuring—e.g. a breathalyzer test, which may be affected by whether there is alcohol in your mouth.

Operational error relates to how we are conducting the measurement—e.g., a problem with the measuring device or how we read it.

While relatively stable in a range, most people's actual level of job satisfaction is inherently unstable.

The same applies to breathalyzer tests. Breath analysis is by far the most commonly used method of testing for blood alcohol (BAC) in impaired driving cases. Assume a police officer takes two separate readings from a driver he has just pulled over. He will likely get different BAC levels between the first and the second reading depending on whether the driver had just burped or vomited; if there was electrical interference from a cell phone and police radio; or if there was tobacco smoke, dirt, or moisture in the environment.

Consequently, from a statistical perspective the BAC level is inherently variable.

To the notion of inherent variability, we can also add operational error. Perhaps the police officer forgot to perform a manual calibration check on the device. The battery was not fully charged. The device was improperly used. The breathalyzer forms were not completed correctly. There was an error in copying down the results, 0.8 instead of 0.08., or between testing the BAC and recording it the officer forgot the actual number.

The point is that, try as we might, it is generally difficult, if not impossible, to have totally accurate measurement. Believing we can do so is simply fooling ourselves. Furthermore, for most situations, “close” is good enough. What does it matter if the BAC is 0.08 or 0.085? One thing that makes statistics powerful is that statistics assumes some error will appear in our measurement.

What is also great about statistics is that, when used appropriately, we can estimate how much error exists in the measurement process.

From the statistician’s perspective, people who believe that total accuracy in measurement is possible are like ostriches with their heads in the sand.

It is far better to accept that error in measurement is everywhere, so why not admit it and try to get an estimate of the size of that error? How can we do that? The answer is that we need to either take several measurements of the same item, or to measure several items assumed to be the same.



From the statistician’s perspective, people who believe that total accuracy of measurement is possible are like ostriches with their heads in the sand.

Descriptive Statistics

Remembering the characteristics of a single item is relatively easy, whether that item is a person, an event like the eclipse of the moon, or the colour of one's motor vehicle. Similarly, most of us can easily recall the characteristics of several items. The more items we have, however, the more difficult it is for us to remember the individual items that make up the group. For example, we may recall the ages of all of our colleagues in the office. Recalling the age of all volunteer staff in a district or municipality is virtually impossible. If we want to be able to say something about the ages of volunteer staff in a region, we need to somehow aggregate or summarize the data. This is where descriptive statistics come into play.

What descriptive statistics do is summarize the characteristics of a group so that we can make sense of a mass of information. Even if we could remember them, listing the ages of 600 men and women is not a very useful exercise. Descriptive statistics allows us to identify certain useful characteristics of the list. Often, the first two things we want to know about a list or bunch of observations are: what is typical, and how much variability is there?

The most common measure of *typicality* is the arithmetic average or the mean. We might find, for example, that the average

Descriptive statistics summarize the characteristics of a group so we can make sense of a mass of information.

We may measure *typicality* by determining the average or median age in the group.

We may measure *variability* by determining the youngest and oldest ages in the group, the spread of ages within the group, or how much the results deviate from the average.

volunteer in a recreation program is 62 years of age. Other measures of typicality include what we call the median and the mode. The median is that point in the age distribution below and above which half of the ages fall. The median age might be 55. In other words, half the volunteers in our program are above age 55 and half are younger. The mode is another term for the most common age. The mean, the median and the mode are the most often used measures of typicality. We can also think of those measures as a central anchor point for the list or distribution of ages.

Measures of *variability* give us an idea of how widely a bunch of measures range or vary. It is one thing to know that the average age of a voluntary staff member in our region is 62; it is something else to know that most are between the ages of 55 and 70 as opposed to between 60 and 65. The most common measures of variability are what we term range statistics and variance statistics.

Range statistics are simple measures of the distance between two points. For example, among our volunteers, the youngest may be 24 and the oldest 72. The range would simply be 72-24, or 48 years. This range measurement is based on the difference between the minimum value in the distribution and the maximum value. Min-max ranges are interesting but can sometimes be misleading. For example, the oldest person in a region might be 78 while most of the other “elderly” volunteers are less than 60. Here, we sometimes call the 78-year-old an outlier.

To deal with distributions that have the odd extreme case, we sometimes use a statistic known as the interquartile range. To get the interquartile range, we need to figure out the age of the person who is at the 25th percentile point of the distribution, and the age of the person who is at the 75th percentile. The interquartile range is simply the difference between those two numbers. Again, like the min-max range, the interquartile range gives us an idea of the spread of the ages.

Besides ranges, we often use statistics known as variability statistics to give us some notion of how the data are spread or disbursed about the measure of central tendency. The two most commonly used variability statistics are the variance and something called the standard deviation. At first sight, these statistics may appear a little daunting but conceptually, they are quite simple. The key in understanding them is not to focus on the math but to consider the underlying ideas.

Examples of typicality and variability are on the following pages.

An Example of Typicality

Even simple descriptive statistics can be useful in decision making. Let's examine the number of parking fines in two neighbouring towns over a week (7 days). The reported fines for each day are shown in the table.

		Town A	Town B
	Day	Number of Parking Fines	
Median Point	Mon	10	9
	Tues	11	16
	Wed	8	7
	Thurs	9	7
	Fri	9	8
	Sat	9	7
	Sun	5	7
Sum		61	61
Mean		9	9
Median		9	7
Mode		9	7

For measures of typicality, we can calculate the average or arithmetic mean, the median and the mode. The average or arithmetic mean is simply the sum of the fines divided by the number of days. The median is that point below and above which 50 per cent of the numbers fall. The mode is the most commonly recorded number of fines.

The data in the boxes represent the actual numbers of parking fines. Even from this limited amount of information, there are several points of interest. First, both towns have a total of 61 fines in a week. This resulted in an average of nine fines a day. Examining the numbers, however, it appears that Town B had one day when there were 16 fines issued. In statistical language, we call exceptional values such as this, outliers. The arithmetic mean is very sensitive to outliers. This is easy to visualize if we replace the 16 with a value of 30. All the other values stay the same but the mean would shoot up to 10.7 fines.

A measure that is much less sensitive to outliers is the median (or midpoint, as it is sometime called). As we have noted, the median is the value that breaks the distribution into the upper and lower 50th percentile. In the table, the median or midpoint is nine, which coincidentally falls on Thursday, the middle day of the week. For Town A, the median or midpoint of the distribution is 9 and for Town B, the median is 7.

That Town B has a lower median than mean is a consequence of the fact that, except for the outlier value of 16 fines, Town B generally has lower numbers of fines than Town A. Because we are only dealing with a few values, this is easy to see. It would be less obvious with a large data set. Regardless, the principles hold.

An Example of Variability

In this example, we will use the parking fine data from the previous box. We have seen that the typical or average occurrences of fines are about the same for both towns. However, looking at the raw data suggests that there might be more variability in the occurrences in Town B as opposed to Town A. The fact that the mean and the median were slightly different provides numerical support for this view.

	Town A		
	Number of Parking Fines	Deviation from Mean	Deviation Squared
Mon	10	1	1
Tues	11	2	4
Wed	8	-1	1
Thurs	9	0	0
Fri	9	0	0
Sat	9	0	0
Sun	5	-4	16
Mean	9	0	2.9
	Town B		
Mon	9	0	0
Tues	16	7	49
Wed	7	-2	4
Thurs	7	-2	4
Fri	8	-1	1
Sat	7	-2	4
Sun	7	-2	4
Mean	9	0	9.4

One measure of variability is the range. Town A's fine rates go from a minimum of 5 to 11, providing a range of 6. Town B's fine rates go from a minimum of 7 to a maximum of 16, providing a range of 9.

Another two commonly used measures of variation are the variance and the standard deviation. While seemingly complex, these measures are conceptually simple. In the second column of numbers, we have subtracted the mean from each individual fine occurrence. For example, in Town A, the first deviation is $10 - 9 = 1$. We do that for each of the individual fine occurrence.

In column three, we simply square the deviations from the means (that is, multiply the value by itself). When we do this for all of the observations, we discover two things. First, the average of the deviations from the mean is zero. This will always be the case because the mean is in the “middle” of the distribution and the positive deviations will cancel out the negative ones. That is why we calculated the third column: the squared deviations.

The mean or average of the squared deviations is known as the variance. The variance for Town A is 2.9 and for Town B it is 9.4. This suggests that there is much more variation in the robbery rates of Town B than for Town A. The variance is a statistic that is used a great deal. In slightly more advanced statistics, our goal is to try to explain why there is more variance or variation in one set of numbers than another. Perhaps in Town B there is a large weekend outdoor farmer's market that attracts large crowds from other cities and towns, which could explain the high fine rates on the weekend. Those are notions or hypotheses we might want to test.

Since squared values generate big numbers, we often compare the square root of the variances. This brings the values back to the size of the original measurement (raw numbers as opposed to squared ones). The square root of the variance is known as the standard deviation. The standard deviation for Town A is 1.7 and for Town B is 3.1. This suggests that the variation in the parking fine rates in Town B is slightly more than twice that of Town A.

An Aside

We can also use variance statistics as an estimate of how much error in measurement exists. For example, two people may take 25 minutes on average to complete an activity. The variance for one person might be eight minutes and for the second person three minutes. Based on the average both people appear equal in performance, but the variance measures suggest that the second person is much more consistent and, in that sense, better. From a management perspective, the interesting question is why one person is more consistent in their performance than the other.

Subsequent investigation may show that the first person has to perform the action under a variety of conditions while the second faces fewer environmental challenges. It may also be that the first person lets things “slide” for a while and then turns on the juice to get the numbers back up to an acceptable average. Regardless, knowing differences in variances can sometimes tell us more than simply knowing differences in averages or central tendency.

Inferential Statistics

The second leg on which the discipline of statistics stands is what we term inferential statistics. Inferential statistics help us to draw conclusions and make decisions. Unlike for most descriptive statistics, the math behind inferential statistics can get complicated. Consequently, we will restrict our focus to the logic underlying inferential statistics and examine how they can be used to help us make decisions. Learning inferential statistics by oneself from a book is typically not easy. For readers who have no background in the

area, it might be worthwhile investing in a one-semester course in a local college. Otherwise, understanding the concepts is sufficient; just leave the details to an expert.

Inferential statistics are used for many purposes. However, the two primary ones are to be able to estimate or infer the characteristics of a population from a sample, and to estimate whether significant differences exist between two or more populations or samples.

Population Estimates

Let's start with the issue of making inferences from samples of populations. If we wanted to know the proportion of the population of a city that uses carbon monoxide (CO) detectors, we could contact each household and pose the question. Collecting information from everyone in a jurisdiction is known as conducting a census. In a city of 300,000 households, that could be an expensive and time-consuming proposition. That is why censuses are done only rarely and under limited circumstances. Fortunately, early in the 20th century, statisticians figured out how to estimate the characteristics of the whole (a population) from a subgroup or sample.

The key to being able to do this, however, is in the way in which the sample is drawn or collected from the population. Essentially, "any old sample" doesn't cut it. The sample has to be taken from the population in a particular way. There are some variations on the theme, but let us keep this simple and consider the basic case. What we want is something statisticians call a simple random sample. A simple random sample is one where each household in the population has an equal chance of being selected, and that chance of being selected is independent of the other selections. Let us break that down into the constituent parts: random selection, equal chance, and independence.

Random selection

Random selection implies the households in the sample are chosen using a chance mechanism—things like coin tosses and computer random number generators. In other words, someone cannot choose the households based on availability or door colour. Random selection implies that a listing of households (say a city directory) exists where the households are listed or numbered from 1 to 300,000. For a sample of 1,200 households, we would use a random number generator to give a listing of 1,200 numbers between 1 and 300,000. Once we have those numbers, we would then identify the households that hold those positions or numbers on the list.

Equal chance

Equal chance implies that each household has the same chance or likelihood of selection. Lists with duplicate addresses or lists that omit a certain type of household (say, all apartments or all households in a particular neighbourhood) mean some households either have a greater likelihood of selection, or no chance of selection.

Independence

This implies that the selection of one household does not determine or affect the selection of another. For example, the person selecting the sample might notice two houses on the same block or two houses next to each other appear on the list. Thinking they might be too much alike, the researcher drops one household in favour of another selection. That is not acceptable. The selections that appear must be included despite anything else.

If we follow these rules, then estimating the characteristics of the entire population from the sample is possible. Some other things need to be considered, such as the size of the sample, but those are details that are best discussed with a professional. If we follow the basic rules outlined above, we can estimate what proportion of the population of households that have carbon monoxide detectors within a given likely range.

In other words, the sample estimate will be close to what actually exists in the population but will probably not be the exact figure. What differentiates statistical sampling from other procedures, however, is that it is possible to estimate the range within which the population figure will likely fall. Thus, we could conclude that the likely proportion of homes with CO detectors we would see is X per cent within plus or minus Y percentage points in, say, 19 surveys out of 20.

The uninitiated often disparage statistical estimates for not being able to provide exact values. But, as we discussed earlier, the fundamental assumption in the world of statistics is that all measurement entails error, so the best we can do is come up with a point estimate and a reasonable notion of its level of accuracy. This is something no other procedure can do. With a statistical estimate, you get an idea of whether an estimate is precise enough to be useful or too variable or inaccurate for practical purposes.

Many other ways of generating estimates are available, but with those, you usually have no way of knowing if they are close to the actual value in the population or somewhere out near the planet Mars.

Significant Differences

Another primary use of inferential statistics is to be able to estimate whether two samples are similar or different. For example, over a year, a Police or Fire Chief might wish to know whether differences in response times exist across stations. Typically, data such as response times are collected through an automated dispatch system. At the end of a period, calculating the mean or average response time is possible. As discussed earlier, the mean value will be an estimate based on error-prone data and there will be a distribution of values around that estimate. Thus, the question is, if the response time of one department is eight minutes and another one is nine, does that one minute difference reflect a real difference or is it simply within the realm of possible measurement error?

Some differences are big and substantively meaningful and do not require statistics to help us make a decision. For example, if the difference in response time were 10 minutes, then we know a real and important difference exists. However, when we get to one minute, it is not clear that the difference is real or just within the realm of normal variability.

What statistics can do is let us know whether that difference is within or outside that range of normal variability. If it is outside, then we say that the difference is statistically significantly different.

We should note, however, that just because something is statistically significantly different, it does not necessarily mean that it is substantively different. For example, the people of Bigtown may earn, on average, \$100 per year more than the people from Smalltown. This difference might be statistically significant but few people would think it is of major importance if the average in both towns was around \$70,000. On the other hand, if something is not statistically

significantly different, then we should assess the difference as being within the normal range of variation and, consequently, not substantively significant either.

Inferential statistics are even more useful when we have multiple comparisons to make. Typically, a large municipality may have hundreds of street intersections. Are the differences in accident rates across all of the intersections significantly different? More advanced techniques can help us to figure out what factors might be related to those differences. That brings us to our final topic in this chapter, and that is the role of statistical modelling.

Statistical Modelling

For most decision makers, the real power of statistics lies in the ability to model social, natural and mechanical processes. Statistical models allow us to examine complex issues where multiple factors might affect a particular outcome. For example, statistical models have been used to study traffic noise on nearby roads. The fact that traffic noise contributes to an area's overall noise pollution is well established. Traffic noise from highways creates problems for surrounding areas, especially when there are high traffic volumes and high speeds. This noise is considered a serious threat to the environmental health by some.¹

In statistical modelling, most of our focus is on trying to explain variation. Thus, we go back to one of our basic statistical concepts—that of the variance. So, for example, we might ask: What are the factors that likely affect traffic noise in different locations or at different times? Is it traffic flow factors, road factors, vehicle factors or human factors?² Based on the outcomes of those and other modelling exercises, it is possible to identify what form of intervention works to reduce noise and what does not work.

Improving the allocation and efficiencies of their resources is also something decision makers might do. Whatever the complexity of the model or underlying process, statistical analyses help us to figure out many useful results with an estimable level of accuracy.

Among the key questions we can address with statistical modelling are the following:

- Does the overall model accurately reflect the process we are trying to describe or emulate? In other words, is it statistically significant?
- How much of the variation in the outcome factor is explained by the model?
- Which elements in the model are statistically significant and which are not?
- What is the relative impact or rank ordering of various components of the model on the outcome factor?
- Are those impacts large enough to be meaningful from a substantive or policy perspective?
- How do the various sub components in the model interact with one another as to their impact on the outcome?

As we indicated, statistics is not the magic bullet for all decision making. Used appropriately, however, statistical techniques can provide a great deal of insight into the questions we are examining.

Decision making is a complex process, and the best processes are those where we use the many tools at our disposal to help come up with an answer. Often, trade-offs have to be made. Something may be statistically significant but not substantively significant. Similarly, just because one choice is more effective than another does not mean that it can be justified socially or economically. Regardless, knowing whether something has a “real” impact or not is a good starting point.

Notes

1. Subramani, T., M. Kavitha and K.P. Sivaraj (2012) “Modelling of traffic noise pollution.” International Journal of Engineering Research and Applications 2: 3175-3182.
2. Subramani, T., M. Kavitha and K.P. Sivaraj (2012)

Experimental Designs

How Do We Know What it Means?

A basic notion underlying this book is that making decisions based on evidence has advantages over other forms of decision making. By evidence, we are referring to observable and measurable “facts” or data. While we argue that it is generally a good thing to have facts, a single fact or bit of data or piece of information is fairly meaningless by itself. The reason for this is that nothing has meaning except in comparison with something else.

Assume, for a moment, that you are on a trip to India and you see a pair of shoes on sale for 2,859 rupees. If you are not familiar with prices in India, you might ask yourself whether this is a good value or not. The “fact” that the shoes are 2,859 rupees is irrelevant to you unless you have something with which to compare it. That comparison might be with another product or with the average hourly wage in India or with the equivalence in another currency. Currently, 2900 rupees is approximately equivalent to \$60 Canadian. It is only by making a comparison that the relative value of the shoes takes on meaning.

Similarly, your local police department might have an overall crime clearance rate of 40 per cent, with a rate of 70 per cent for violent offences.

To understand the meaning of a fact, we need an appropriate point of comparison.

At a city council meeting, the question is raised as to whether these are acceptable performance rates. The average citizen might have expectations that at least 90 per cent of all crimes result in charges being laid or being otherwise cleared. By referring to national data reported to Statistics Canada, it can be shown that the overall clearance rate in Canada is about 60 per cent for violent crimes and about 40 per cent for crime overall. By making this comparison, it is clear that your department is performing on par with the rest of the country for overall crime, and somewhat better when addressing violent crime.

The point being made is that to understand the meaning of a fact, we need an appropriate point of comparison.

Within the framework of evidence-based decision making, a key question we have to ask ourselves is: What is the most appropriate point of comparison?

A complementary question might also be: What is the best way in which to make that comparison? The answer is to use a standard framework that program evaluators and applied scientists call experimental designs. Experimental designs are simply different approaches to helping us make an appropriate comparison.

The remainder of this chapter will focus on some basic experimental designs that we use to assess the value of information or data related to a question about which we need to make a decision. In applied research, designs can become very complex. No matter the complexities of the design, however, there are a few fundamental principles that underlie the value or the merits of the design.

The “Counterfactual”

When we do or observe something, the question is: What would have happened if the event had not occurred? What if the Axis powers had won World War II? What if the party in power had not won the last election? What would have happened if insurance companies provided fire services instead of municipalities?

The comparison is with some theoretical model. It cannot give us proof of something, but as a mental exercise, it forces us to identify the important elements of a policy or program. What are the relevant or active components that are making the difference or that we expect to have an impact? These ideas, which are counter to the existing outcomes or facts, are called “counterfactuals.”

Albert Einstein referred to this mulling of counterfactuals as thought experiments. Thought experiments

consist of conducting an analysis in our heads to think through the potential impacts and consequences of a particular event or outcome. What differentiated Einstein’s thought experiment from simple fantasizing or theorizing is that he also focused on how we might test the thought experiment using real situations and observable data.

As an example of a thought experiment, we might consider the issue that employee performance is affected by the level of stress caused by the nature the job content; for example, the perceived risk, long hours, shift work and level of responsibility and accountability.

We recognize that these elements can be stressors, but do they in fact affect one’s level of performance? In our thought experiment we might consider other factors such as organizational stressors.

Experience tells us that other factors can affect job stress levels. Perhaps it may not be the nature of the work that generates the greatest amount of stress for our staff. Instead, it is the characteristics of the organization and behaviours of the people in them that may produce stress. Maybe it is the lack of rewards or recognition for a job well done that is affecting the job performance. We should also consider other job-context factors that are likely to create stress in the office, such as organizational structure and various aspects of organizational life (for example, co-worker relations, training, resources, leadership and supervision). Through this thought experiment we conclude that job content is not the sole causal link to job stress levels, but that other stressors such as job context are strong contributors.

By thinking it through, we have come to a conclusion that makes sense. In itself, though, what makes sense logically does not always work out in the observable world. What we need is hard evidence based on repeatable observations—evidence that lies not just in our heads but evidence that can be seen, shared and evaluated by others.

What Makes Up Good Evidence?

When we engage in evidence-based decision making, the fundamental question is: What makes up appropriate evidence? If we think of science as a mechanism for finding the “real” explanation of

What makes sense logically does not always work out in the observable world. What we need is hard evidence based on repeatable observations.

something, then thinking of it within the context of a court case makes sense. In the courts, as in science, there are varying amounts of evidence provided.

Even if something is fundamentally true, we perceive some evidence as more valid, more reliable and more relevant than others. So it is in science. Good evidence stands up to the rigours of a good cross examination. Still, what makes up good evidence?

One characteristic of good evidence is how rigorously people have tested it. Within the framework of science, the basic mechanism for testing an idea is the experimental design. Experimental designs are physical applications of logic, so let us examine the logic underlying experimental designs.

Assume for a moment that we wish to assess the impact of burglar alarms on home break-in rates. One approach would be to take a community and install burglar alarms in all residential homes. We could then see if a difference existed between the break-in rates before and after the introduction of the alarms.

The inability to observe the same entity in two different situations simultaneously is known as the “counterfactual problem.”

Unfortunately, any difference might be the result of other factors (recall our previous discussion of spuriousness). For example, by coincidence, home break-in rates might have dropped due to a decrease in the number of young people in a neighbourhood, or a more positive job market, changes in police patrolling, or perhaps due to a more active neighborhood watch program. We know all of those factors are related highly to home break-in incidences.

Ideally, we would like to be able to observe the same community with and without burglar alarms simultaneously. In other words, we would assess the effect of a burglar alarm program based on the difference in outcomes for the same community with and without participation in the program. Nevertheless, we know that this is impossible. Something cannot be in two states at the same time. At any moment the community either participated in the program or did not participate.

The inability to observe the same entity in two different situations simultaneously is known as “the counterfactual problem.” That is, how do we measure what would have happened if the other situation had existed?

If we cannot assess what would have happened if the opposite or counterfactual situation occurred, then how can we decide if burglar alarms have an impact and not something else? The approach scientists and program evaluators take is to find a comparison group that is as close to the treatment group as possible. How close that comparison group is to the treatment or experimental group determines how much credibility we can have in our results.

There are many ways of finding or creating comparison groups, some of which are better than others. The adequacy of a comparison group is something evaluators spend much time and energy considering.

For example, we might find a “sister” community not far from the target community and use that as a comparison. On the other hand, we might decide to hand out burglar alarms to every second residence, or to residences on the south side of the community but not on the north side. We might even consider comparing our target community with all of the other communities in the province or region.

All of those approaches can provide a point of comparison against which we can judge the potential impact of burglar alarms in the target community.

The problem, however, is that all of those options have possible limitations. Some conditions or circumstances make the target and the comparison group inherently different. Sometimes we can see those differences. For example, in selecting a “sister” community, it may be that the residences in that town are older and tend to have a poorer overall security design. That might be an obvious difference, even to a casual observer. Often, however, the differences are not obvious.

The remainder of this chapter will focus on the different ways we might identify valid comparison groups to accurately reproduce or mimic the counterfactual. Identifying such comparison groups is the crux of any impact evaluation, no matter what type of program we are evaluating. Simply put, without a valid estimate of the counterfactual, we cannot establish the impact of a program with any degree of certainty.

Comparisons With Targets (The One-shot Test)

One of the simplest designs we have is to compare our population of interest with a particular goal or standard. Often, policy guidelines are based on legislated standards or targets set from studies of best practices. Targets can vary according to the context. For example, a community might target a 20 per cent reduction in traffic accident incidents over a five-year period. A parts manufacturer may implement a six-sigma regime, where one expects that fewer than 3.4 defective parts per million will be manufactured. Human resource policy may also dictate that organizations should strive to hire a certain percentage of individuals belonging to minority groups.

The key, then, is to compare our population of interest with a target that is theoretically doable or achievable. Once we implement

an action, the question becomes whether we have met the target or goal. If we achieve the target, we have reason to believe that the action (which is generally a policy or program implementation) has been successful. Of course, we will use a statistical procedure to help us determine whether we are close enough to the target to be equal to the target.

The methodological literature sometimes calls this approach the one-shot test. That is, an action, policy or program is carried out, compared with a standard and, if it meets the standard, we generally assume the action was successful. The evidence might seem reasonably convincing. Unfortunately, one-shot tests have their limitations. We can see one major limitation in the following example.

The one-shot test does not account for alternate explanations for a result.

Example: One-shot Test

Suppose a community has a fire death rate of nine per million population and wishes to reduce it to five per million over a three-year period. The Fire Chief might decide that handing out free smoke alarm is the most cost-efficient way of achieving this goal. He carries out the program and three years later, the death rate is 5.1 per million which, given the size of the community, is statistically equivalent to the target of five per million. Can we infer that the smoke alarm program is behind the reduction in fatalities? The evidence seems compelling.

In fact, an alternate explanation for the reduction might exist. The free smoke alarm campaign generated substantial publicity in the local press. Firefighters and volunteers went door to door distributing the smoke alarms. A notice left at the door asked citizens not at home to pick them up at various retail outlets. Together, the campaign generated substantial awareness of issues relating to residential fire safety. Because of the publicity, people in the community became more aware of the need for fire safety and made other changes in their homes. Some cleared clutter from around furnaces, fewer people used space heaters after going to bed, and more people planned escape routes should fire occur in their houses.

In other words, by heightening awareness of domestic fires, the community members took actions that would have reduced the likelihood of fatalities regardless of whether they had installed the smoke alarms.

The point here is not to argue that smoke alarms do not work in reducing fatalities. The point is that there may be alternate or coincidental explanations why the target was met. How much credibility those alternate explanations might have depends on different factors. First, does it make sense logically that the alternate explanations might hold? If previous publicity campaigns resulted in no noticeable impact then we might wish to stick with the smoke alarms as the effective mechanism. On the other hand, if publicity campaigns in other communities had resulted in substantial drops in death rates, we might be more supportive of the alternative explanation. A further explanation might be that fire death rates were declining overall for a variety of reasons, such as longer-term changes in building code, overall heightened awareness, decreases in smoking rates, and so on. Consequently, the death rate would have declined regardless.

Before-and-after Designs

A variation on the one-shot or target design is the before-and-after design. Again, we have a group or community of interest where we are looking to make an impact. We measure the situation beforehand, apply some intervention and then look at the outcome later. The assumption here is that any difference between the after and before results is due to the impact of the intervention. Unlike the one-shot design where the comparison is a policy goal or target, the implicit comparison in this design is the after results with the before baseline.

The before-and-after design shares most of the strengths and weaknesses of the one-shot design. Specifically, we can never be sure if it is the intervention that had an impact or simply some coincidental effect. For example, a jurisdiction might want to reduce the automobile accident rate among young drivers. The way they decide to do this is by dropping the legal Blood Alcohol Concentration (BAC) limit from .08 to .05 for drivers under the age of 25. Examining the data from the three years before the introduction of the legislation with the data from three years after, an evaluator notices that accident rates have indeed dropped for younger drivers.

Again, we might consider the change in legislation to be the precipitating factor. On the other hand, it is possible that rates of drinking and BAC levels among young drivers have not changed.

The difference is simply due to the increased vigilance of the police, who are targeting younger drivers in an attempt to enforce the new legislation. It is likely similar police vigilance without the change in legislation would have produced similar results. That is, the important factor is not the legislation, but simply enhanced surveillance by the police that serves to act as a general deterrent to young drivers.

Looking Past the Limitations

The limitations of these designs do not mean the evidence collected is irrelevant. We would have good reason to believe the results if we impose these interventions in many communities and under different circumstances with similar outcomes. Also, carrying out an intervention and then revoking it can tell us a lot. If the intervention results in the desired outcome and the revocation results in a return to the original baseline, then we have a more powerful argument that the intervention is the causal factor. What we need to remember is that evidence is rarely absolute. It has varying degrees of reliability or credibility associated with it. Just as in the courts, some forms of evidence are more credible than others.

Given the inherent weaknesses of these designs, we might ask what approaches we can take to address the problem. So far, the gold standard among evaluators and scientists is what we term the classical experimental design.

The Classical Design

A rule of thumb in science is that nothing is perfect and certainty is an elusive goal. On the other hand, a lack of certainty in one's death is rarely a reason for playing Russian roulette. Similarly, a one per cent risk that one will lose all of one's assets in the stock market generally results in a different form of investment behaviour than if the risk is above 80 per cent. So, if we do not have perfection, what is the current ideal or gold standard for experimental designs?

To date, evaluators and scientists have relied on the two-group, before-and-after design to provide the most valid and the most reliable evidence. We start with the before-and-after design mentioned above. We then complement it with a comparison or control group that serves as the counterfactual. In other words, we have one group exposed to a treatment and one group that is not. If the group exposed to the treatment exhibits a significant change and the comparison group does not, then we have very strong reasons for believing the intervention had an impact.

The key to the strength of classical design experiments is to ensure the comparison (control) group is equivalent to the experimental group.

The key to the strength of this design is to ensure the comparison group is equivalent to the experimental or treatment group from the outset. This harkens back to our earlier discussion of the counterfactual where, ideally, we would like to see the same elements exposed to the treatment and not exposed simultaneously. This situation is physically impossible. However, we can ensure that both the treatment and comparison groups are initially as alike as possible. How do we do this?

One way is to take pairs of identical people (or communities or what have you), and divide them into two groups. However, unless the pairs are exact clones, we can never be certain that they are identical on all relevant characteristics. Fortunately, while we can rarely work with clones or identical matches, we can divide subjects into two statistically equivalent groups. As we have noted previously, statistically equivalent does not mean truly identical, but it does mean that, on average, no statistically significant difference exists between the two groups. In other words, for all practical purposes, they are close enough to being identical.

The method for ensuring statistical equivalence is to take an initial group and randomly assign the members to the treatment and the comparison groups.

By random assignment, we mean using something like a coin flip (with a fair coin) or a random number generator to make the assignment. With a large enough initial group, the resulting two sub groups will be statistically equivalent. That is to say, any significant differences among individuals across the groups will cancel themselves out. To a point, the larger the initial group, the more equivalent the two sub groups will appear.

Any systematic factors that might affect the outcome (beyond the intervention) will be distributed across the two groups. Thus, the two sub groups will be the same on all relevant characteristics, except that one is exposed to the intervention or treatment and the other is not.

Avoiding Sample Selection Bias

The key to having a strong classical design is for the researcher to conduct the random assignment to the experimental and comparison or control groups. Situations where we have not randomly assigned subjects to treatment and comparison groups have the potential for what we call sample selection bias. What this means is that the treatment and comparison groups might differ on a relevant factor. For example, we might conduct a study of residences that have burglar alarms with those that do not.

If crime rates are lower in residences where the residents have installed burglar alarms, it may not be that most or all of the difference in the lower crime rates is due to the burglar alarms. It is quite possible that people who install burglar alarms are more conscientious than people who chose not to do so. In other words, those who installed alarms are also the same people who have taken care to install high quality locks or window bars, and are active volunteers in the Neighbourhood Watch.

Usually, any situation where people or subjects volunteer or select into the treatment group should be considered suspect. Subjects often volunteer for a program because they are more motivated or see the treatment as potentially more beneficial. Sample selection bias can only be addressed if the evaluator or researcher has done a random assignment to the treatment and control conditions. Having said this, it is imperative that the researcher engages in true random assignment. It is not unknown for some researchers to select those they think will be the most cooperative or most likely to succeed to be in the treatment as opposed to the comparison group.

Less Than Ideal Variations

Sometimes we cannot randomly assign members of a group to policy or program intervention and others to the control. A situation where this often arises is when governments decide to legislate policy. By their nature, social policies are implemented throughout a jurisdiction and not randomly assigned within particular areas. What happens, for example, if the Province of British Columbia wishes to introduce a new set of performance standards regarding hospital wait times? Obviously, we can apply the before-and-after model, but we know that has limitations. Are there ways of using the framework of the classical design to overcome those limitations?

Matched Comparison Designs

The answer is, some approaches are less ideal than the classical model but perhaps more convincing than simply using the before-after approach. Since we have no ability to randomly assign jurisdictions to different response standards, one approach is to find potential clones. That is, jurisdictions with different standards that we know (or, more likely assume) to be similar in all or most relevant aspects. For British Columbia, we might consider choosing Washington and Oregon States, and the Province of Alberta as comparators.

The assumption here, of course, is that these jurisdictions have different response standards but have similar geographical and socio-demographic characteristics to British Columbia.

We call this approach the matched comparisons procedure. We attempt to find matching jurisdictions that are as similar as possible to the experimental one(s) to provide a relevant control group. Again, the issue of sample selection bias might arise, since there is likely something different about jurisdictions that decide to implement a policy over those that do not. Just as with the simple before-and-after approach, we need to regard these results with greater suspicion than those obtained from the gold standard of the classical design.

Regardless, matched comparison designs have produced convincing evidence that certain practices are effective. Perhaps one of the best examples is the early research into the use of daytime running lights on automobiles for reducing traffic accidents.¹ On the flip side, matched comparison studies have also suggested that some policies do not have the intended impact. A good example here is the research into the relationship between capital punishment and homicide rates.

The preponderance of the cross-jurisdictional evidence suggests that while capital punishment may assuage our feelings for revenge, it does little to reduce actual amount of homicide.

We need to make a decision and the stronger the evidence, the more likely the decision will be the correct one. We could be wrong, but even wrong decisions help us know what does not work. Doing the same thing over and over makes no sense if the results do not change. When it becomes obvious that our current practices do not have the desired impact, logic suggests we should try something different. Eventually, we are likely to find something that does work. An important factor is that we must be willing to change our view when faced with contrary evidence.

Doing the same thing over and over makes no sense if the results do not change.

Too often, we ritualistically engage in the same behaviour even when the evidence shows it doesn't generate the outcome we wish. For centuries, physicians engaged in bloodletting because, despite the evidence, it seemed to make "common sense" at the time. The fact that many patients were unnecessarily weakened by the practice and subsequently died, was not a consideration.

The Essentials

The important point behind this discussion is that how evidence is collected—the framework or design used to generate the data—is an important element in helping us determine how credible the evidence might be. Among the key factors is our prior notion that nothing has any meaning unless it is in comparison with something else.

In other words, everything needs a comparator for us to be able to make sense of it. An intervention or an action only makes sense in comparison with another action or a non-action (doing nothing). That comparator is known as the counterfactual.

Since something cannot be in two different situations at once, we must look for the most appropriate comparison. As we have seen, clones are hard to come by, so the best approach we have devised to date is the randomized experiment where subjects or objects of interest are randomly assigned to a treatment group and an appropriate comparison or control group. The randomization process helps ensure that there will be no systematic sample selection bias.

In some cases, random allocation to treatment and comparison group is not possible, so we try to create situations that come as close to that ideal as possible.

Evidence generated by these approaches should always be considered suspect but,

if the approach appears sound and there are few logical alternative explanations for the effect, then we are generally willing to give the evidence reasonable weight until we find something superior.

Even with the best designed experiments, however, the results are not always equally credible. The design is one element we consider; the magnitude of the impact or size of the effect being produced is another factor. Obviously, interventions that produce large effects provide better reasons for using the evidence for a decision than small or marginal effects. But that leads us to other considerations such as policy evaluation and cost-benefit or cost-effectiveness analyses—topics of our next chapters.

Notes

1. See, for example, Elvik, R. (1993) "The effects on accidents of compulsory use of daytime running lights for cars in Norway" *Accident Analysis and Prevention*, 25: 383-398.

Program Evaluation

Introduction

All levels of governments spend most of their annual budgets delivering services—public safety services, fire services, environmental protection services, social services, transportation services, health services, parks services, maintenance services, and more. Examining the different jobs of government, you will find most are associated with the delivery of services. It is not surprising then, that governments everywhere are trying to determine whether or not they are best meeting the needs of the people they serve. Accordingly, governments regularly re-examine levels of service to ensure they are adequate and appropriately targeted. They will also assess whether services are structured and operating in the most effective and efficient manner possible. All of this is to ensure taxpayer dollars are well spent.

Assessing the efficiency and effectiveness of service delivery is not simple. Things can get complicated very quickly. One of the primary issues is that governments rarely have sufficient resources to meet service demand. Further, when governments want to make changes to service delivery, they are commonly faced with the constraints of infrastructure shortcomings, labour agreements, jurisdictional concerns, legislative requirements, and many underlying political pressures.

Assessing the efficiency and effectiveness of service delivery is not simple.

This is why cutting, changing, or adding services is always a difficult exercise. The result is that there is a significant difference between what governments wish they could or should do, and what they actually can do. Consequently, evaluating services delivered by government is a sensitive issue and it is little wonder that governments are often wary of evaluations, especially when they are not placed in context.

Evaluating services becomes even more sensitive when one considers that many services provided by all levels of government are delivered by external organizations. They are delivered by businesses, independent contractors, and a multiplicity of non-profit or not-for-profit service agencies. In reality, many of these agencies do not have the resources or in-house expertise to adequately evaluate the services they provide. Moreover, there is an inherent problem with doing self-evaluations because most organizations have a vested interest in presenting themselves in the most positive light possible.

On the other hand, governments cannot afford to do thorough evaluations on the services provided by every organization they fund. Typically, funding arrangements are often for very short periods and the amount awarded is often limited. In short, while governments might wish they could conduct evaluations in such cases, resource constraints inhibit them. Accordingly, governments often rely on an individual's or organization's reputation, and take at face value the worth for services provided by those they contract. Fortunately, almost all external government contracts are limited and contractors know they have to maintain basic standards in order to have their funding renewed.

Having said this, room remains for government agencies to assess the impact and value of many internally and externally delivered programs. Evaluations do not have to be complicated, expensive or laborious. They can also be done with respect for the sensitivities all governments must consider when they assess the services they provide.

With those constraints in mind, the first thing is to recognize that all government services can be thought of as programs of one kind or another. They may be called initiatives, social enterprises, pilot projects, courses, or just plain services, but we can look at all as programs that can be evaluated as self-standing entities. All are supposed to deliver a product or service in a way that something is accomplished.

Furthermore, those accomplishments are supposed to be implemented in the most efficient way possible. In an ideal world, we could also compare programs of interest against alternatives and determine which are superior. From this perspective, what we are talking about is a single technique called program evaluation. Knowing the basics of program evaluation will help you know what to look for when assessing whether or not a service is effective and gives taxpayers good value.

This chapter will review what questions to ask in assessing a program. Clearly, some programs are so large and multi-jurisdictional (for example, some United Nations initiatives) that evaluating them requires a background beyond what we can provide here. Similarly, some programs are so multi-faceted in their purpose and outcome that the methods needed to evaluate them are exceedingly complex. However, the approaches we will address are appropriate for assessing most of the “bread and butter” programs governments deliver.

At the end of the chapter we will discuss program logic models to help guide you through the evaluation process. First, though, we need to get a handle on the basic questions that should be considered before starting an evaluation.

The First Question: What is the Program *Exactly*?

The key word here is “exactly” because unless you know the details of a program being provided, you cannot really measure its full effect, and you certainly cannot determine whether or not it operates in an efficient and effective manner. Moreover, you cannot ensure you are comparing the program to its appropriate alternatives because you may be unwittingly comparing apples with oranges.

Having said that, we find that this first question is rarely asked – people often assume that once a general program description is provided that is sufficient. This is not good enough. You need to know enough details about the components of the program so that there is no mistaking what is being delivered. An organization might state, for example, that they are offering a restorative justice program in a community. This is fine as far as it goes, but there are many different varieties of such programs and the differences among them are such that you would be hard pressed to find two alike once you determine what they actually do.

A program description must always include a clear articulation of what people receiving the program are expected to receive. Often, you will know you have a good description when the components of the program are defined unambiguously and are measurable.

A program description must always include a clear articulation of what people receiving the program are expected to receive.

Without this, it is impossible to get a good answer to the next question to be asked in a program evaluation. Regardless, the importance of having a well-articulated description of what a program entails will become clearer as we consider the evaluation process more fully.

Perhaps a good way to consider the point is to think of a weight-loss regimen. You need to describe what that program looks like in a way that allows outsiders to measure what the participants are expected to do and receive. As we all know, weight-loss programs can be of varying lengths and take many forms with many component parts (e.g., diet, exercise, trainers, and supplements), and many look deceptively similar at first glance. Accordingly, a general program description is not enough.

The Second Question: Did the Program Deliver What it Was Supposed to Deliver?

Again, this is a question that is rarely asked. Yet, it is typically not one that is difficult to answer if the program is set up initially to document how program delivery takes place. Those receiving the program, for example, can be asked if they actually received each aspect of what it purported to offer. They can be asked how much of each element of the program they received. They can be asked if they were even involved in the program.

To get an appreciation of the point here, you need only think back to your high school or university days when you took a particular course. You will recall that not all courses were as described in the course outline, and just because there was a teacher in the classroom did not mean the course material was covered in a way that students actually learned something. Moreover, even when the material being delivered was as planned, not everyone enrolled actually participated. Some students slept through the course, some were daydreaming, and some were simply absent. Commonly, great differences appear in student evaluations of the same university course taught by different professors.

Some students indicated the course offered less than it should have; for example, a required textbook was never referred to, exam questions had nothing to do with the lectures, lectures had nothing to do with the course outline, or the professor was hard to understand.

If this happens when we are talking about a simple program such as a university course that has been offered for years for a fairly homogeneous group of students in a fairly defined setting, you can imagine how program delivery can vary when a program is offered in a multiplicity of settings, by a multiplicity of service providers to a broader range of recipients. Again, one simply cannot assume that the program was delivered as expected or that it was received as intended. To know what is really going on, you need to audit claims of delivery which will include measures of delivery.

To reiterate, the point is that just because someone was in the program does not mean that they involved themselves as prescribed, or that they got access to component parts as intended. This second question requires that you have a way of confirming the extent to which participants received and completed the program as prescribed.

Both pre- and post-measures are required to identify the degree to change.

The Third Question: Did the Recipients of the Program Actually Benefit from it?

Having satisfied yourself that you know the exact nature of the program and the extent to which it delivered what it was supposed to have delivered, you should be ready to move to the ultimate issue: outcomes. The key here is establishing pre- and post-measurements to determine the extent to which the recipients of the program (e.g., cities, organizations, neighbourhoods, targeted groups, and individuals) experienced a change in something (e.g., conditions, satisfaction levels, attitudes, skills, capacity). That change should relate back to whatever it is that the program was specifically intended to make happen.

Here, pre-measures are extremely important. These provide an indication of where program recipients are starting, thus giving you a base of comparison for whatever influence the program might provide. This also respects the fact that not all recipients are starting at the same level.

Normally, a discussion of the pre-measures to be chosen will be a consequence of available data and what

indicators are tied directly to the post-measures. Without these pre- and post-measurements, you have no way of knowing whether the program had the intended effect. That said, if you choose your pre- and post-measures thoughtfully, you can likely determine what aspects work best for which participants, when and where, and under what conditions.

To help put the matter of pre- and post-measures in perspective, let us consider a Block Watch crime prevention program which works from the premise that if neighbours know each other better, are attentive to the homes of neighbours, report suspicious activity, and do a number of things to better safeguard their own homes, crime will decrease. But, the first part of the program evaluation should not concern itself with whether or not crime goes down. We first need to confirm that we have answered the first question that we are actually talking about Block Watch with all its components. That is, did the implementation include neighbourhood meetings, the printed materials and a Block Watch Captain to organize neighbours to keep them informed?

Next, as indicated by the second question, we need to know how many of the neighbours participated in each component of the program. That is, did they attend meetings? Did they make a point of getting to know their neighbours? And, did they follow home security recommendations and lock their doors and windows as recommended?

Once we have confirmed that neighbours were invited to participate in the full Block Watch program, and that they actually participated, we need to address the third question to determine whether or not Block Watch caused neighbours to do what they weren't doing before the program, and if they did, to what extent they did those things.

A pre-measure, at the start of the program, might include asking targeted neighbours how many of the neighbours living beside and across from them they have actually talked to. It might also include asking neighbours about what steps they had taken to protect their home and property. If this seems to be going a bit far, we know of one study that looked at the effectiveness of Block Watch and determined that nearly everything that the program was intending to do was already being done by homeowners in surrounding neighbourhoods not involved in Block Watch.

That study didn't include pre-measures, only post-measures. It is a safe bet that if the analysis had included pre- and post-measures, it would be revealed that the program had not really changed anything with respect to participant behaviour. Meanwhile, the city involved with the program had been paying a staff member full-time to coordinate the program—clearly a waste of tax dollars.

To emphasize the point using the weight-loss program, clearly we would want to know the weight of participants both when they entered the program and when they completed it to see how much, if any, weight they had lost. Ideally we would have other background information on participants to indicate for what type of person the program worked best. We would also want to be sure that the program was directed at people who needed to lose weight in the first instance and were not already doing other things to lose weight.

The Fourth Question: Was an Ultimate Benefit Achieved?

More often than not, the “ultimate benefit” question gets confused with the third question which asks whether or not participants or their conditions changed because of the program. Again, we can look at the Block Watch program to get a sense of the difference between questions three and four. In the case of question three, we are trying to establish whether or not the participants actually changed their behaviours as a consequence of being part of the program. This would indicate whether the program is working or not as intended. The next question takes us to the overriding purpose of the program, which in the case of Block Watch, is to prevent or reduce crime.

Importantly, this “ultimate benefit” question is not one you can simply jump to without addressing question three because many things could be influencing the ultimate goal. That is, you might never know whether it was the program influencing the ultimate goal or something else. We might, for example, determine that a program is working as intended but, in the end, it does not significantly impact its ultimate goal. In the case of Block Watch, the study also found that the crime rates in Block Watch neighbourhoods were the same as in comparable and surrounding non-Block Watch neighbourhoods. As mentioned, we already know from addressing question three that the Block Watch

program, as rolled out in at that particular instance, was not accomplishing what it was supposed to accomplish, so we should not have expected it to make any difference in crime rates.

On this matter of assessing ultimate benefit, it is important to have a comparison group or situation so one can determine what might have happened without the program being in place. Programs sometimes appear to be effective in accomplishing an ultimate benefit when that benefit is occurring elsewhere because of factors that have nothing to do with the program. This is certainly the case with many crime prevention programs that claim to be effective, but have essentially ignored the fact that crime rates have been dropping almost everywhere in the Western world.

In any case, it is one thing to confirm that a program is doing what it is supposed to be doing, as asked by question three, but it is something else to confirm that it is contributing to some ultimate goal. This requires two separate analyses, involving two sets of pre- and post-measures and, ideally, two sets of comparison groups – one relating to each of questions three and four.

To reiterate the point, we can consider the issue of the weight-loss program.

Question three requires that we at least measure the extent to which participants were successful in losing weight, while question four requires us to measure the extent to which losing weight contributes to some other overriding programmatic goal such as better health. This latter consideration could be operationalized

in a number of different ways, such as looking for less overall illness, fewer trips to the doctor, fewer sick days taken, or fewer medications dispensed. Another way to look at it is that losing weight is only important if it actually contributes to making something else happen.

The Fifth Question: So What?

Whenever an evaluation is completed, one should ask whether or not there is another program that can do what the evaluated program was intended to do but more effectively. Even if one determines that the evaluated program is meeting expectations, one should still be looking to see if an even better mousetrap exists. But, that is only the first part: you also need to ask if there is another program that would accomplish the ultimate goal more effectively. Accordingly, you need to compare your results with those of other programs.

Making comparisons can be done in a number of ways, but a good start is to review the literature on the subject area relating to the program. The literature is full of reports on evaluations of programs and, with a little effort, you are likely to find information pointing to what has been determined to work and not work elsewhere. With luck, you might even find a meta-analysis which will show you how

a collection of programs like the one you evaluated compare. Care needs to be taken to ensure that you are not comparing apples to oranges. Ideally, you will do a literature search before you start the evaluation, and in the process discover how others have conducted similar evaluations.

With this in mind, the weight-loss program is a good example. It may be that the program helped people lose weight, but there may be other programs that can achieve the results more effectively. Bearing in mind the ultimate goal of broader health outcomes, perhaps other programs can accomplish those goals more effectively, for example, with simple diet changes or some lifestyle alterations. Regardless, the literature is full of examples of both weight loss and other programs designed to improve peoples' health in one way or another. The goal is simply to ensure that the program being evaluated can be determined to be among the best ways of achieving the ultimate goal.

The Sixth Question: Is the Program Operating as Efficiently as it Could?

Until now, our focus has been on what can be technically described as “outcome evaluations.” That is, we have been focusing on establishing whether or not a program is doing what it is supposed to be doing (the intermediate outcome), and on whether or not it is contributing as expected to a broader goal (the ultimate outcome). An equally important part of program evaluation, however, is what we refer to as “process evaluation.” Process evaluation is an exercise in assessing the step-by-step operations and systems associated with a program to examine whether it is implemented in the most efficient manner possible. Accordingly, it involves taking an in-depth look at the resources being used, assessing them in amount, quality, and application, and determining whether or not they are best for what the program needs. Sometimes programs are under-resourced in both human and financial terms. Sometimes they are over-resourced in one way or another. And, sometimes, programs need a re-alignment of resources. It may also be that resources are simply mismanaged.

The content of a program may also need revision. Leadership, intake procedures, referral systems, data systems, technology, accountability mechanisms, communication issues, labour matters, and stakeholder involvement, may also need to be examined.

These need to be done with the goal of ensuring that all of the tasks associated with a program are being carried out in a way that best provides what the program needs to deliver its outcomes.

The importance of doing a process evaluation cannot be emphasized enough. All of us have gone through programs that do not operate as they claim to do. It is easy to be misled about a program’s potential because of a hidden weakness in implementation. Every good program also stands a chance of being better if a process evaluation can identify operational improvements. We need to remember that effectiveness is at risk when a program is not running efficiently.

A process evaluation can often seem threatening to those involved in running a program. But, it does not have to be. Not every aspect of the program has to be placed under a research microscope. The evaluation can start in a general fashion with attention to the most relevant tasks and systems or those with issues or concerns.

Process evaluation assesses the step-by-step operations and systems of a program to examine whether it is implemented in the most efficient manner possible.

You can always look further where concerns exist, assuming time and resources permit. Further, the evaluation can be carried out as a “formative evaluation,” where the overriding goal is to come up with recommendations for improving efficiency and effectiveness.

Looking again at the weight-loss program, we can see that rather than focusing on outcomes, a process evaluation would

likely examine how the program is being managed and administered. This might include looking at the ways participant involvement is tracked, and the ways in which participants access the diet, exercise, and supplemental program elements. The goal would most likely be to generate recommendations on how to make the program run in a more participant-friendly and efficient manner.

The Seventh Question: Does the Program Represent Good Value for Money Spent?

We have already noted that we should ask if there is a better mechanism to achieve the ultimate goal of the program. We should also be asking whether there is an alternate program that can do the same thing at a lower cost. At a cursory level, this is a straightforward exercise: one establishes the costs of the program and then looks at the cost of competing or alternate choices. At a more detailed level, the exercise commonly requires considerable experience and skill, especially once you start trying to factor in indirect costs, contributions in kind, multiplier effects, and the like. In any case, it all falls under the umbrella of cost-benefit or cost-effectiveness analysis as discussed later in this book.

Costing analysis is not just about comparing the cost of one program to another. It may also involve addressing the question of

whether the program is saving resources as expected. Programs are often put in place with an expectation that they represent a less expensive way of doing something. That is, they are intended to represent a cost savings in the first instance.

Giving attention to cost analysis in the weight-loss program scenario, we might want to know, for example, whether or not the program is less costly than similar programs. We also want to know whether the overall health benefits gained through any weight loss actually represent a cost saving over the investment in the program. We might even go so far as to look at whether there are other, more cost-effective ways to achieve whatever health benefits are accrued through the program. Again, the goal is to ensure that the program represents good value for the resources invested.

A Way to Organize Your Evaluation: Using a Logic Model

Thus far, we have discussed evaluations in terms of some fundamental questions. At the same time, however, those questions can be used as the basis for a “logic model” – a framework to help guide the assessor through the evaluation process. Logic models can have different levels of complexity. You can get a sense of what might be involved by considering the following.

1. Program Activities – Here, as in the first question, the specific activities designed to generate each of the program’s intended direct outputs or results need to be identified. Accordingly, you should consider the techniques applied, the products and technology used, and the strategies of how the program functions to produce each expected output. For example, if the program being evaluated was a life skill program, you would need to know such things as what curriculum was being used; the method of delivery; how many hours of instruction were involved; the qualifications of the facilitator or instructors; the delivery format; the delivery schedule; and, what materials

were being used. Typically, you will know you have a good description when an informed outsider is able to understand the program without having seen it. An informed outsider should also have a good appreciation of how and why the activities are related to the intended outputs or results.

2. Outputs – Consistent with the second question, the point here is to confirm that the program delivered what it was supposed to, in the amounts and quality described. In the case of the life skills program, for example, you would want confirmation of the extent to which the format was followed, which materials were used, which aspects of the curriculum were delivered, and the extent to which participants had an opportunity to receive the knowledge and skills presented in the course. Another way to look at this, is that while *program activities* is about auditing the intended components of a program, *outputs* is about measuring and auditing whether the program was delivered as intended.

3. Immediate outcomes – Here we focus on question three and look for confirmation that the program produced a benefit for its recipients. Basically, this is an exercise in measuring any change that might have occurred for recipients because of their participation. In the case of the life skills program, for example, this would involve measuring by way of pre- and post-testing whether or not the participants acquired skills, knowledge, behaviours, and attitudes respecting each particular set of life skills that they did not have going into the program. A more sophisticated assessment might include how much they retained from the program at later dates. Further, if the participants' background information was collected, it would be possible to relate that information to participant learning.

4. Ultimate outcomes – As indicated by question four, a key assessment goal is to confirm that the program resulted in some intended ultimate benefit. Again, immediate outcomes are not in and of themselves the reason programs are put in place – they commonly have some broader intended goal. This involves measuring the extent to which the program influenced that goal. Doing so requires a comparison of recipients of the program to non-participants. In the life skills program,

for example, the ultimate goals of the program might be to enhance the employability of groups of offenders, to improve home stability, to reduce substance abuse, or to reduce recidivism. The task then would be to measure whether or not, over some follow-up period, offenders participating in the program had higher rates of employment, better home stability, reduced substance abuse, and lower rates of recidivism, than did a group of similar offenders who had not participated in a life skills program.

5. Comparison outcomes – Here, as in question five, the task is to determine whether or not there is a better alternative out there. In this regard, there may be versions of the program implemented elsewhere that could serve as good comparisons, or published results on alternative programs may be available in the literature. Regardless, one needs to be mindful of the results of alternatives to assess whether the program under evaluation is truly a best option. It is not sufficient for the program to meet its ultimate goal if an alternative can meet those goals more effectively. Using the example of the life skills program, the task with respect to comparing outcomes would be to do a literature scan of the results and impact of other life skills programs.

6. Activity efficiency – Question six raised the matter of whether the program is operating as efficiently as it could be. Consequently, one would do a review of the resources used and operational procedures with a view to determining whether the stated outputs could be achieved in a more efficient manner. With the life skills program, one would likely be looking for whether the program needed to be as long as prescribed, whether or not materials and class time were fully used, and whether or not course size could be increased without hurting program effectiveness.

7. Cost-benefit comparison – Question seven points to completing a cost-benefit analysis to address two issues. First, is there an alternative program or path to the ultimate goal that represents better value for dollars invested? Second, what is the cost of the program relative to the cost associated with not having it in place? In the instance of the life skills program, for example, this would involve establishing its costs and then comparing those to an alternative life skills program or the cost of not having a program at all. In other words, is the cost of running the program more or less than the costs associated with the amount of crime that non-participants generate?

The process is outlined in the accompanying chart.

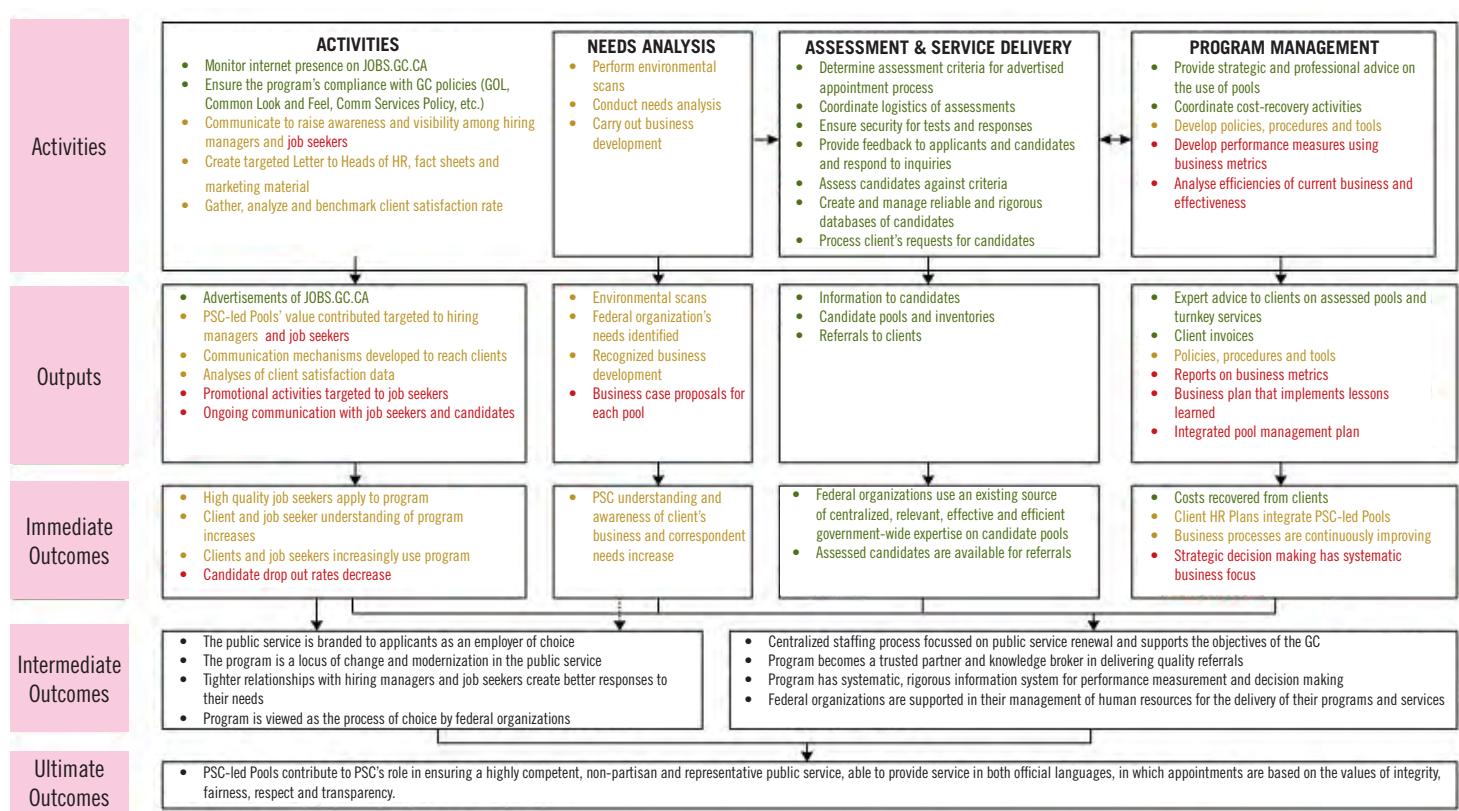


Example: Public Service Commission Logic Model

The logic model¹ below is a visual representation of the inputs, activities, outputs and outcomes of an initiative. This one in particular was done by the Public Service Commission of Canada (PSC) to analyze and identify strengths and weaknesses of PSC-led pools based on the projected goals. PSC-led pools are a new and innovative way for the PSC to fulfill its role as a common service

provider and to complement other PSC services, namely staffing, assessment and other pools of pre-tested candidates, such as the Post-Secondary Recruitment Program.²

They are listed by activity stream. As PSC-led pools are fairly new and still evolving, the operational team is learning while doing and trying to minimize the gaps between how PSC-led pools should operate and how they actually function.



Based on the exercise on the previous page, the following gaps were identified:³

Communications and outreach

The biggest gap found in this stream is in the relationship between job seekers and PSC-led pools. Survey results suggest that job seekers have limited awareness and understanding of the procedures for PSC-led pools. Candidates surveyed felt that the PSC did not keep them well informed of their status in a PSC-led pool (66 per cent).

Needs analysis

Environmental scans, needs analysis and recognition of business development opportunities must be started in some regions and formalized in others. Business case proposals for each pool have to be developed systematically. At the moment, these activities are conducted in an ad hoc fashion.

Assessment and service delivery

Since service delivery is core to PSC-led pools, the operational team has focused most of its efforts and resources in that area. However, there are still some gaps in how activities are carried out, particularly in providing feedback to job seekers and candidates. These activities seem to be the strongest area of PSC-led pools.

Management of the initiative

Business metrics and other data sources, such as management information processes, are key tools for assessing and measuring performance and results. As of November 2009, performance measurement data range from limited to inadequate, and standardized national procedures do not exist. This situation creates complexity in assessing success.

The framework we have outlined is not the only one you can use. A quick search of the literature will lead you to a number of others. The key, however, is to have a systematic framework for examining what a program is designed to do. Also, there may be reasons why a program evaluation does not refer to each component discussed here. What we have presented is only a guide.

Summary

As stated at the beginning of this chapter, evaluating government-funded services can be complicated. But it helps when you view them as programs to be evaluated. In fact, we would argue that most government services can be seen in this way, and assessed under the umbrella of program evaluation. Further, we see the exercise of program evaluation as one where the evaluator begins with a set of foundational questions in mind as we have posed here. This is not to say that every evaluation will involve addressing each question. Still, if the goal is to assess whether or not a service being delivered actually works as intended, that it is working efficiently, and that it represent a good financial investment, each question needs to be considered.

The questions we have presented here are only the beginning. For each of them the real work is in developing a research design that enables you to get an answer that is evidence-based and with which you can be confident. Accordingly, that involves a consideration of the other issues that we cover in this book. As any experienced researcher will tell you, one rarely gets to do an evaluation as comprehensively as one might want. Many things typically get in the way such as a lack of data, inaccessibility to detailed program information, time and budget constraints, and other factors you cannot control. The goal, though, is to be as rigorous as circumstances allow, carefully calling attention to whatever limits and cautions need reference in the description and presentation of the results.

Notes

1. Public Service Commission of Canada
<http://www.psc-cfp.gc.ca/abt-aps/inev-evin/2010/pools-bassins/img/figure4-eng.jpg>
2. <http://www.psc-cfp.gc.ca/abt-aps/inev-evin/2010/pools-bassins/index-eng.htm#ex-sum>
3. <http://www.psc-cfp.gc.ca/abt-aps/inev-evin/2010/pools-bassins/index-eng.htm#appC>

Costing Analysis

Basic Concepts

Costing analysis comes in one of two variations. The first instance deals with the costs associated with doing something. For example, the decision to purchase a vehicle involves not only the capital cost of that vehicle, but also maintenance such as the cost of repairs, consumables such as gasoline, and support costs such as insurance. Depending on the circumstances, additional support costs may arise, such as those associated with having to build a new garage or rent a parking space. If we are looking at the true cost of ownership, we should also factor the depreciation of the vehicle (hopefully, we will recuperate some capital cost when we sell it in a few years) plus the interest on the funds used to purchase the vehicle.

The other form of costing analysis is what we term a cost-benefit or cost-effectiveness analysis. In this instance, we weigh the costs associated with the decision with the value of the expected benefits. For example, a department might choose to invest in further training. The question then arises: What is the return on that investment? If the training relates to how to fight online crime in a community where the internet does not exist, the return on investment might be considered zero. In fact, it is a straightforward cost situation.

Costing studies allow us to identify the total cost of a decision and the associated benefits.

On the other hand, if the training relates to staff health and safety matters, the returns may appear in lower accident and injury rates, fewer sick days, lower insurance rates, more efficient or productive employees and higher employee morale. We can weigh the relative value of those benefits against the cost associated with the training sessions to estimate the relative return on investment.

A fundamental idea of economics is the notion of opportunity cost. Assuming you have a limited budget, deciding to do one thing necessarily precludes another. For example, given a department's capital budget, the decision is made to purchase a pick-up truck. By making that choice, the alternatives—an SUV, a sedan, a motorcycle, and so on—are foregone. That is to say, the opportunity to select an alternative is no longer available. Not only is the physical choice of the next best alternative not available, we give up the benefits associated with that choice.

Costing studies help us to identify the total cost of a decision and what the returns or benefits associated with that decision might be. Furthermore, we can also examine what we consider the expected cost and returns associated with the second or third best choices, and compare those to our preferred decision. Sometimes this exercise results in our seeing a “lesser” alternative as superior to our initial preference.

Monetary costs are not, nor should they be, the only factors that we consider when we make a choice. Political and other social considerations influence how we make choices. However, monetary costs are important and are typically easy to quantify. Most products and services have a monetary or market cost associated with them. Also, social and political costs are often closely linked to economic decisions. As with formally assessing monetary costs, using the general costing framework to assess the impacts of non-monetary decisions is also possible. The only difference is that in those situations, the costs and returns are often more difficult to quantify. Regardless, decision makers can and do use qualitative data to weigh the impact of those types of decisions.

No matter whether we do a straight costing analysis, cost-effectiveness or cost-benefit analysis, there are five overall steps to consider.

Steps to Consider

1. Identifying the component in the department's operating or strategic plan to which the question or analysis relates.
2. Setting out the objectives that we intend the decision to achieve.
3. Identifying the options or choices that are available.
4. Conducting a financial (cost-benefit or cost-effectiveness) analysis of the option selected or the options under consideration.
5. Preparing an accounting statement summarizing the results.

These steps may appear to be a restatement of what we have mentioned previously. This is the case. However, we need to see effective evidence-based decision making as part of a broad framework that starts with a consideration of what we are doing and why, what are the alternatives, and what evidence can we bring to bear to help us make a decision. Unless we know what we are doing and why, it is almost impossible to identify the appropriate information. Without knowing that, we may collect much data but we likely will not be collecting much evidence.

Cost Analysis

Straight costing studies involve estimating the total life cycle cost of a particular piece of equipment or service. By life cycle, we are referring to the period during which we use the product or service. For example, a motor vehicle might have an actual average life expectancy of about 12 years before it is ready for the scrapyard. A person or an organization might decide to buy a vehicle, keep it for five years and then sell it. In that instance, for the owner, the vehicle's life cycle is five years.

The key to conducting accurate cost analyses is to ensure that we include all of the appropriate costs. Generally, for equipment or capital goods, these fall into the following categories:

- depreciation,
- interest on capital,
- maintenance fees (consumables and repairs),
- licensing or regulatory costs, and
- operator costs.

While analysts will often exclude operating costs from the analysis, those need to be considered, even if the final decision is to exclude them. If the equipment is meant as a replacement component, then the operating costs would carry over from the previous piece of equipment. However, suppose a municipality has decided to purchase a new fleet of salt trucks and to include a road grader in its inventory.

The key to conducting accurate cost analyses is to ensure that we include all of the appropriate costs.

That additional vehicle may require extra operating and maintenance personnel, the cost of whom we need to factor into the analysis.

Some of you may wonder why we have just included depreciation in our list of items instead of the initial capital cost. Here the assumption is that the piece of equipment will be sold at the end of the life cycle. Consequently, the capital cost component here is the difference between the purchase price and the selling price. This is what we call depreciation.

Different pieces of equipment depreciate at different rates, but it is common for that to be about 20-30 per cent per year. We calculate depreciation on the outstanding value, so a \$100,000 piece of equipment that depreciates at a rate of 20 per cent per year would be worth \$80,000 after the first year. The second year's depreciation would be $\$80,000 \times .2$, or \$16,000. Thus, the total depreciation after two years would be $\$20,000 + \$16,000$, or \$36,000, and the residual value of the equipment would be $\$100,000 - \$36,000$, or \$64,000.

One item often forgotten in costing studies is the interest on the purchase. Interest rates are sometimes called discount rates in the literature. The need to consider interest is generally obvious when one borrows the money to make the purchase, since the bank or financing company will include that charge. However, even where the equipment is purchased outright, we should include the “rental” value of the capital. The reason for this is that if we had not made the purchase, we could have invested the money for a given return or used it for some other purpose. This, in effect, is another form of opportunity cost.

Obtaining Reliable Cost Estimates

Whether it is the total cost of hiring someone or purchasing a piece of equipment, the key to good costing studies is to ensure we include all items, and obtain the most accurate and reliable cost estimates of those items. Because organizations work in different environments, typically we gain the best information from experience. Looking back over your organization’s financial records can be revealing. Because they reflect actual experiences, it is easy to see where unexpected costs (and savings) arose. Do not write those off as unique or one-time occurrences; put those in as line items in your analysis.

Where drawing on institutional experience is not possible, one can often obtain information from other sources.

Often, suppliers will give cost comparisons with competitors’ products. Beware, however, that those analyses often selectively include or exclude “inconvenient” line items. Make sure that you are comparing the proverbial apples with apples. Where you find missing items, ask for supplemental information.

Many independent agencies also conduct costing analyses of various items. Look especially to professional or trade associations. Similarly, non-governmental organizations and other public agencies will often make their budgets and costing studies available. Much of that can be found online or in a local library. Sometimes a simple phone call can result in a gold mine of data.

An example of a straight costing study is presented in the box on the next page. Here, we are looking at the cost of owning and operating a typical, full-size pick-up truck over a five-year period. The cost of the operator is not included in this example.

Straight costing studies are done to estimate life cycle costs to decide the affordability of a purchase. They are also useful in comparing different products. For example, one brand of pick-up might have a higher capital cost but lower maintenance costs than another. The question then becomes: Which is the better choice?

The Cost of Purchasing a New Pick-up Truck for Personal Use

Three-year cost of purchasing and operating a pick-up truck:

Item	Cost (\$)
Purchase price	23,500
Selling price	9,500
Depreciation	18,577
Financing	3,387
Fuel	10,079
Insurance	3,471
Taxes and licensing fees	3,650
Maintenance	2,069
Repairs	821
Total cost	42,054
Cost per kilometre	0.47

Assumptions:

- 20,000 km driven per year
- 2.7 per cent APR financing cost with \$2,750.30 down payment
- gas \$1.25/l.
- mileage at 10.46 l/100 km.

Similar analyses can be used to decide whether it is less costly overall to purchase a used vehicle as opposed to new, or to lease as opposed to purchasing outright. Obviously, for these different scenarios, we must make different assumptions regarding expected life cycle, operating costs and depreciation. It might also be worth repeating that the values used in costing studies are generally estimates. As we discuss in the chapter on statistics, all values are estimates. The key, with a little research and experience, is to minimize the error. However, many expected items, such as the selling price of the vehicle and the actual cost of operation, are based on assumptions that are out of one's control.

We have considered the cost of capital goods but we can conduct similar analyses for personnel. The same general principles apply. Typically, we focus on a person's salary when deciding to hire someone, but ancillary costs can be substantial. When pensions, taxes, insurance, benefits and other compensation-related issues are considered, it is common for those to add an additional 15-30 per cent to the total salary cost. This is above the cost of training and maintaining the person. Maintenance costs include the person's working space and any equipment and supplies they may need to do their job. In the previous example, we noted that equipment typically needs an operator. So, too, people often need equipment to do their job.

A Note on Cost-effectiveness

In the previous analysis, our attention was on the total cost of owning and operating a vehicle over its life cycle. Knowing the total cost of something is an important consideration in decision making. Often, however, knowing the total cost does not tell us the whole story. Most equipment or other items generate some form of output or product. For a car, the output is transportation. In that instance, knowing the cost per kilometre is often a more valuable piece of information than the total cost.

In the example provided on the previous page, the expected cost of the vehicle per kilometre is about \$0.47. We term the price or cost of something per unit of output as its cost-effectiveness. While cost-effectiveness is clearly related to total cost, we should treat it as an independent issue for decision making. Often, differences in total costs might be irrelevant. It is the per-unit cost that is important. One reason unit costs differ from total costs is the fact that total costs consist of two components: fixed or sunk costs, and variable costs. Fixed costs are associated with such things as the one-time cost of purchase. Variable costs generally relate to operating and maintenance costs. A piece of equipment may have a higher fixed cost but, if it is more efficient than a lower priced piece, it will generally have lower unit costs.

A key element in cost-effectiveness analyses is being able to identify the appropriate output measures and being able to measure them appropriately.

The same applies to personnel costs. Higher salaries to people who are more productive, who are less likely to miss work and who provide a better quality of service can outweigh “savings” accrued by outsourcing to lower-cost jurisdictions. What is important is how many items are produced, how many people are served, and the quality of that output or service.

A key element in cost-effectiveness analyses, however, is being able to identify the appropriate output measures and being able to measure them in the right manner. Again, this is where examining the organization’s operating or strategic plans becomes important. It is in those documents that the organization’s objectives and operational purpose should be outlined. Either directly or indirectly, an organization’s effectiveness is related to the product or service it is meant to deliver.

Cost-benefit Analysis

Cost-benefit analyses are generally extensions of simple cost-effectiveness studies. A primary difference is that cost-benefit analyses look at a broader range of returns on the investment. Most cost-benefit analyses include effects (benefits) that are not easily quantifiable or outcomes that have a broader social impact.

Cost-benefit analysis is grounded in welfare economics. It differs from most branches of economics since the focus is not just on decisions of consumers and firms, but on public decisions that affect the economic interests of a broader community. Consequently, cost-benefit analyses often focus on issues such as quality of life or quality of the environment. A fundamental challenge for those doing cost-benefit analyses is how to measure the benefits so they are comparable across issues. Among commodities, apples are not electrical transformers. However, a market for both exists and it is possible to place a monetary value on both. Currency is a common exchange unit that allows the producers of apples to purchase transformers even when the producers of transformers have no interest in exchanging their product for apples.

The difficulty with many public goods and services is that there is no open marketplace in which the monetary value of those items is established. Moreover, for ideological reasons, many people refuse to assume a monetary value on public goods.

A common refrain, for example, is that, “You can’t put a price on the environment” or, “You can’t put a price on a human life.” The fact is, we do both. The problem is that no independent or indifferent market exists to set those prices. Regardless, this is an essential weakness of cost-benefit as opposed to straight costing analyses.

Revealed and Stated Preferences

While the philosophical issue of whether you can truly value a human life may not be answerable, welfare economists have two broad tools at their disposal. They term one approach the revealed preference method. Revealed preferences relate to how people actually behave when confronted by a qualitative phenomenon. For example, comparing a particular piece of real estate with similar ones could reveal the “eyesore value” of having a fire hydrant on a front lawn. How much parents value education for their children might be suggested by what proportion of their income they are willing to spend on a child’s tuition.

The second tool in the economist’s repertoire is what we call stated preferences. Stated preferences are just that: what someone is willing to tell you they would pay for something.

We may judge people's value of environmental elements, for example, by how much of a tax increase they are willing to support for clean air or nature conservatory initiatives. Typically, stated preferences are determined through surveys and similar procedures.

While both stated and revealed preferences have their merits, both have their limitations. Using how much life insurance a person has to assess how much they value their lives might appear like an excellent revealed preference. However, how much they can buy is limited by how much insurance they can afford. Furthermore, a person may value their life highly but not be willing to see relatives "benefit" from their death since life insurance goes to the beneficiary and not the insured. Stated preferences on various aspects have been studied extensively by sociologists for the past century. Their overwhelming conclusion is that what people say and what they do varies considerably.

Still, cost-benefit analysis is one of the few techniques we have to assess the broader impact of various policies and programs. It helps us to clarify the issues, identify the constituent components, and bring some evidence to bear on the issue. It has gained general acceptance in the public sector and is mandatory in many government shops. For example, the Treasury Board of Canada has mandated that any regulatory framework put in place by the federal government must be based on a cost-benefit analysis.

We can use cost-benefit analysis to:

- decide whether a proposed project should be undertaken
- decide whether an existing project should be continued
- choose between alternative projects

The purpose is for "departments and agencies [to] assess regulatory and non-regulatory options to maximize net benefits to society as a whole. Hence, all regulatory departments and agencies are expected to show that the recommended option maximizes the net economic, environmental, and social benefits to Canadians, business, and government over time more than any other type of regulatory or non-regulatory action."¹

In summary, we can use cost-benefit analysis in various ways. For example, to:

- decide whether a proposed project or program should be undertaken;
- decide whether an existing project or program should be continued; or,
- choose between alternative projects or programs.

Components of a Cost-benefit Analysis

In setting up and executing a cost-benefit analysis, several steps need to be followed. These include:

1. Define the problem

Again, this is a statement of the issue with a link back to your operational or strategic plan.

2. Identify any constraints or limiting factors

This is a discussion of what administrative requirements and other challenges you might face. These include a listing of financial limitations, managerial or personnel challenges, environmental and other regulations, and any other factors or “hurdles” you might need to address.

3. List the alternatives

Every initiative has alternatives, including doing nothing or staying the course. For example, if the issue is whether to close a particular office location or not, it may be informative to look at amalgamating with another department, sharing space with other services, or expanding the operation to incorporate other functions.

4. List the benefits

For the alternatives outlined, what is the return on investment? Is there a monetary return or an increase in productivity or effectiveness? Perhaps the matter is not one of generating further revenues, but one of reducing or avoiding costs. Are there health, safety or environmental benefits to be gained? The issue might be related to overall quality of life. Are there savings to be had in equipment, time or personnel?

5. How are the costs and benefits to be quantified?

Clearly, market or monetary values of goods and services are the easiest with which to work. We have already outlined the challenge of providing market values. Still, finding a shadow or proxy price for a given cost or benefit may be possible. Social scientists have developed ways to estimate the value of a human life.² The cost associated with noise levels or high traffic volume in a community, for example, can be estimated by differences in housing values between noisy and quiet communities or between those with high and low traffic volumes.

Often, we can find ways of assessing the value of tough-to-monetize issues by searching the appropriate literature. We have already discussed techniques for conducting more focused online searches. Using the expertise of economists and other social scientists in local colleges and universities might also be possible.

Once we have conducted these steps, we can put a report together summarizing these elements and presenting the relative costs and benefits.

Net Present Value

As the saying goes, “A bird in the hand is worth two in the bush.” So it is with money. One reason we charge interest on borrowed money is that by giving capital to a borrower, the lender faces an opportunity cost. That money cannot be used for anything else. To compensate the lender for the opportunity cost, borrowers must pay interest. For example, when you buy a locked-in savings certificate with a five-year redemption, you get back more than you invested. A \$1,000 certificate invested at three per cent would be worth $\$1,000 \times 1.03 \times 1.03 \times 1.03 \times 1.03 \times 1.03 = \$1,000 \times 1.03^5 = \$1,159$.

We can also consider the opposite. What would an endowment of \$2,000 that you are to receive in five years be worth to you today? In other words, what would you be willing to pay for the benefit of having the cash right now?

This is the principle behind reverse mortgages. A bank or financial institution will give you a fraction of your home’s value today if you allow them to sell it at market value and keep the proceeds several years hence. This is the opposite of the previous problem. In these instances, we call the interest rate the discount rate. At a three per cent discount rate, that future \$2,000 endowment would be worth: $\$2,000 \times 1/1.035 = \$2000 \times .863 = \$1,725$ today.

We term this current value on a future amount its net present value or NPV. The NPV is the opposite of the future value. Since programs and capital goods have an expected life cycle, it is common to standardize costs to today’s value, that is, the NPV. Another way of thinking about NPVs is to consider them as equivalent to constant as opposed to real dollars when we are trying to control prices for inflation.

In these examples, we have discussed what economists call the private time preference rate, since the focus is on an individual. Within the public sphere, the choice to invest public funds in a particular program often precludes investments in other programs of benefit to the public. Within the public or welfare sphere, economists generally call the deferred value the social opportunity cost. While the terminology differs, the underlying principles are similar.

Benefit-Cost Ratios

For programs extended over time, we need to amortize both cost and benefits. Occasionally, the duration of the costs may be different from the duration or life expectancy of the benefits. An extreme example here is the pyramids. The Great Pyramid of Giza was built around 2550 BC and presumably paid for at the time. The Egyptian tourist industry, however, has been reaping the benefits ever since. Consequently, to make things comparable, analysts calculate the NPV of both costs and benefits.

We term the ratio of the benefits to costs as the benefit-cost ratio or BCR. Assuming the NPV of the benefits of

a program is \$13.5 million and the net present value of the costs is \$10 million, the BCR would be:

$$\text{BCR} = \frac{\text{(NPV Benefits)}}{\text{(NPV Costs)}} = \frac{13.5}{10.0} = 1.35$$

Ideally, the BCR should be greater than one. Anything less assumes that the costs outweigh the benefits and, all other things being equal, the option should not be chosen. If we chose to evaluate several alternatives, the one with the highest BCR would normally be our choice. If a program with a lesser benefit-cost ratio is selected, then it is likely that we should have included the reason for that selection on the benefit side of the ledger.

Example: Glenmore Reservoir Diversion, Calgary, Alberta

In June 2013, the City of Calgary experienced major flooding within those parts of the city adjacent to the Bow River.³ This was an unforeseen event that came about because water flowing along that stretch of the Bow River exceeded the once-in-100-year limit. In fact, similar flow levels had not been experienced since the 1930s. Overall, it was estimated that the damage caused by the flood was in the range of \$445 million. An additional \$55 million was allocated to emergency response items.⁴

After the event, the City hired a consulting firm to estimate the cost of constructing a Glenmore Reservoir Diversion Tunnel near Heritage Drive. This is not an unusual civil engineering project and it did not seem to provide inordinate challenges from the outset. The example, however, provides some insight into the different components that go into this type of project. The consulting company provided the following cost estimates.⁵

Continued on next page

Example: Glenmore Reservoir Diversion (*cont.*)

The primary costs (in \$millions), as one might imagine, are associated with the construction of the diversion.

FLOW CASE

Capital Cost Components	500m ³ /s	700m ³ /s
Mobilization	32.7	32.3
Inlet	55.5	63.5
Tunnelling	132.6	146.6
Outlet	68.2	71.7
Other	0.9	0.9
Subtotal	289.9	315
Contingency	72.5	78.8
Total	362.4	393.8

Here, the engineers provided alternative scenarios based on expected maximum water flow volumes of 500 and 700 cubic metres per second. There are a few points to note. First, the costs are estimates (based on values of mid-2014) and the actual amounts would likely vary once the contract went to tender and actual labour and material costs were calculated. This, plus the fact that there may be unforeseen challenges that might arise or changes made by the city to the specifications, results in the “contingency” item listed just below the subtotal. As is common practice with these types of contracts, the contingency fee is set at about 25 per cent of the total estimated capital cost.

Another item omitted from the cost estimate is the Goods and Services (value-added) Tax that might be incurred.

The estimated capital cost of the project is not the only one that would be borne, however. Surveying, engineering, right of way and other costs also add to the total. Including those items, the consulting engineers provided the following total estimated cost for the project.

FLOW CASE

Total Estimated Costs	500m ³ /s	700m ³ /s
Capital Costs (Construction)	362.4	393.8
Environmental Mitigation	5.4	5.9
Professional Services	90.6	98.4
Right of Way	0.1	0.1
Total	458.5	498.2

In this latter table, we see that after construction costs, the next largest item consists of “professional services.” These include construction management fees, design fees, permits and other items that are a standard part of any large project. Again, the consultants used a rule of thumb that professional services typically come in at about 25 per cent of the capital construction costs. The right of way entry is the cost of a construction easement that would be necessary during the construction stage. Once again, value-added taxes were not included nor, for that matter, were expected lifetime maintenance costs for the diversion.

Continued on next page

Example: Glenmore Reservoir Diversion (*cont.*)

The net costs for this diversion would be in the range of half a billion dollars which is in line with the total estimated costs of the 2013 flood.

The looming question is whether this investment is worthwhile? Most likely, the affected home and business owners would

agree. Others in less susceptible areas of the city might have differing views. Needless to say, the question engenders a debate over what is the likelihood of another event of this magnitude in the near future, and what are the acceptable policy alternatives?⁶

Example: Public Transit Fleet Purchase, New Zealand

Another pertinent example is the decision faced by municipalities regarding what type of bus to purchase for their municipal transport fleet. Several factors fit in here including the purchase price and environmental considerations.

Nunns, Varghese and Adli looked at some options for the basis of a future public transit fleet in New Zealand.⁷ The standard vehicle they considered was based on diesel technology. While less expensive than gasoline, diesel fuel poses some challenges. Diesel engines have significant emissions and they can be noisy. Alternate technologies are available including diesel/electric hybrid models and fully electric vehicles. Nunns and his

colleagues noted that, at face value, diesel buses were the least costly to purchase. The cost per vehicle for a standard diesel powered vehicle was in the \$300,000NZ to \$450,000NZ range.

Diesel/electric vehicles cost about \$600,000NZ each and fully electric vehicles were in the \$900,000NZ to \$1,000,000NZ range. From a capital cost perspective, traditional diesel-powered vehicles seemed like the obvious choice. After doing some “whole-of-life” cost projections, however, the cost differences started to look quite different. The following table summarizes the team’s findings regarding alternate fleet costs.

Continued on next page

Example: Public Transit Fleet Purchase (*cont.*)

Fleet scenario	Bus purchase (incl. fixed costs)	Fuel	Bus maintenance and renewal	Driver	Total
Better diesel buses	\$165.9M	\$184.7M	\$277.7M	\$385.2M	\$1013.5M
Hybrid bus introduction	\$220.1M	\$159.9M	\$276.3M	\$385.2M	\$1041.5M
Diesel then electric	\$256.1M	\$159.1M	\$286.2M	\$385.2M	\$1085.5M

Across an expected 12.5-year average life span of a fleet, there appeared to be little difference in the overall costs associated with the type of drive system. Consequently, it would appear that decision makers might want to focus on ancillary factors such as the level of emissions or noise, or the proven reliability of the technology. As the authors note,

the hybrid and fully electric buses gave off fewer emissions and were generally quieter. However, their technology was less tested over the long term and performance was sometimes an issue. The fact that there was less infrastructure to support the newer technologies was also an issue that needed consideration.

Summary

While costing studies are but one way of generating data for evidence-based decision making, they are often one of the more commonly used tools. Essentially, costing studies do three things for us. First, when done properly, they link the outcomes we wish to measure with the goals and objectives of our operational and strategic plans. They essentially help us focus on the question about whether the activity is within the organization's mandate.

Second, costing studies help us to focus on the many line items that make up actual costs. Often, "back of the envelope" or convention-based costs omit many ancillary costs associated with our activities. For example, it is common for costing studies to omit interest payments or costs associated with the need for extra personnel. By focusing on a detailed analysis, we are more likely to ensure that we include those items. Furthermore, exhibiting the results of a costing analysis to colleagues and others provides the opportunity for independent observers to identify potentially missed items.

Third, costing studies provide a transparent and fairly mechanical way of helping us decide on options.

The assessments are relatively objective and focused. The assumptions underlying the costs can be scrutinized, as can the values associated with individual items. The transparency of the process provides for a more defensible decision: one that is replicable by an independent observer. Furthermore, unlike purely value-based decisions, decisions based on evidence force critics to generate alternate values or analyses to validly criticize the analysis presented.

Even if someone can put forward alternate evidence, a net benefit still exists since that evidence will contribute to a more accurate assessment of the situation. In the end, a better basis for a decision is put forward.

Notes

1. Treasury Board of Canada (2007) *Canadian Cost-Benefit Analysis Guide: Regulatory Proposals*. Ottawa: Government of Canada. Catalogue No. BT58-5/2007.
<http://www.tbs-sct.gc.ca/rtrap-parfa/analys/analys-eng.pdf>
2. Robinson, L.A., (2007) "How US government agencies value mortality risk reductions." *Review of Environmental Economics and Policy*, 1: 238-299.
3. For a report outlining the extent of the disaster, see Expert Management Panel on River Flood Mitigation (2014) Calgary's Flood Resilient Future. Available at: <http://www.calgary.ca/UEP/Water/Documents/Water-Documents/Flood-Panel-Documents/Expert-Management-Panel-Report-to-Council.PDF>
4. The report of the Expert Panel estimated that the total cost of flooding in Province of Alberta that year was in the range of \$5-6 billion. Only a fraction of that was covered by insurance.
5. See http://www.calgary.ca/UEP/Water/Documents/Water-Documents/Flood-Panel-Documents/Appendix_G_Cost.pdf ; more information on the project is available at the City of Calgary website at: <http://www.calgary.ca/UEP/Water/Pages/Flooding-and-sewer-back-ups/Flood-Mitigation-Panel/Flood-panel.aspx> [as at August 6, 2015].
6. For a nice summary of technical articles relating to disasters, see: Shreve, C.M. and I. Kelman (2014) "Does mitigation save? Reviewing cost-benefit analyses of disaster risk reduction." *International Journal of Disaster Risk Reduction*, 10: 213-135.
7. Nunns, P., J. Varghese and S. Adli (2015) "Better bus fleets for New Zealand: Evaluating costs and trade-offs." Presented at the IPENZ Transportation Group Conference, Christchurch, New Zealand, March 22-14. Available at: <http://conf.hardingconsultants.co.nz/workspace/uploads/paper-nunns-peter-better-54f39398eebe2.pdf>

Making Decisions

Using Evidence

We make decisions all the time in our private and professional lives. Mostly, those decisions are based on what we learned in our training, on conventional wisdom, or on traditional practices. Often, questioning common practice only leads to rediscovering the wheel. Yet, there are many circumstances where traditional practice and common knowledge do not work. We may not achieve the results we want, or our practices lead to less-than-efficient outcomes. For some reason, however, humans are reluctant to change. We are a conservative species. We become comfortable doing the same thing repeatedly, even when we are not happy with the outcome. As the Alcoholics Anonymous Handbook states, however, “Insanity is doing the same thing, over and over again, but expecting different results.”

Evidence-based decision making makes the process transparent—it is no longer a closed, magical process, but one where observers can follow the logic and follow the evidence.

Historically, we can forgive decision makers for pursuing timeworn rituals. After all, as rainmakers knew, if you danced often enough, it would eventually rain. Modern weather forecasting has become sufficiently accurate, however, that rainmaking is no longer a viable profession.

The reason for that is meteorology has accumulated sufficient systematic knowledge that it is possible to predict local temperatures, precipitation and other phenomena with a high degree of certainty. Meteorologists have accomplished this by turning to scientific research and other forms of systematic study.

The reliance on systematic study and data collection, which is what underlies science, has made inconsistent inroads in most other disciplines.



This is unfortunate since, today, there is a large amount of empirical evidence to help us make better decisions. Furthermore, where existing analyses do not exist, conducting a local analysis to improve our own decision making is often not that difficult. This doesn't mean that one needs to become a scientist—far from it. All we need to do is to use empirical results to be able to build a reliable body of evidence.

Decision making based on evidence will generally allow you to make better decisions. Evidence-based decision making has the advantage of making the process transparent. Outsiders can become privy to the foundations of the decision. It is no longer a closed, magical process but one where observers can follow the logic and follow the evidence.

Evidence-based decision making is using the best available research and information on the outcomes of government policies and services to carry out guidelines and evaluate agencies, departments, and personnel.

We are not suggesting that you can always find an optimal solution to your problem. However, evidence-based decision making helps us to identify options and practices that do not work. In those instances, you are likely no worse off trying something new. Most often, however, a review of the existing evidence or the collection of your own data will help provide a more fruitful direction.

Everyone draws inferences from evidence. Inferential reasoning is a basic human skill. Thinking analytically is a skill like drawing and painting or operating a vehicle. It can be taught, it can be learned, and it can improve with practice. However, like many other skills such as karate, it needs to be hands-on and applied. This manual, companion workbook and related case studies will afford you that opportunity.

In summary, how can we put the lessons of this book together to formulate a good evidence-based strategy for decision making? Essentially, there are four main steps.

Without the right question, no amount of data will help provide an answer.

Identify and Frame the Question

The first three chapters of this book are focused on identifying appropriate questions. Without the right question, no amount of data will help provide an answer. We have stressed repeatedly that good questions need to be put into an appropriate framework. Ideally, you should draw these from your organizational plan or your strategic plan. This helps to focus the issue on the key purpose and objective of your unit. One main reason many organizations fail is that they lose sight of their mandate. They try to be all things to all people. This is simply not achievable.

If you lack an organizational or strategic plan, the next best thing is to drill into the issue. Ask several fundamental questions:

- Why are we proposing to do this?
- What are the likely outcomes?
- How does this action relate to the organization's mission?
- What benefits will this action bring to my organization or the people we serve?

- Are there more cost-effective or cost-efficient alternatives?
- Does this action have long-term or short-term consequences?
- What other resources am I likely to need if we pursue this action?

If what you are proposing to do is new or outside the traditional scope of your organization's mandate, consider putting together a focused business plan to support or justify the activity.

Once you have identified and justified the appropriate question, outline the options.

Commonly, two or three viable alternatives are available. In other situations, the range of options and their relative merits is not necessarily obvious. In those situations, consider performing an environmental scan or SWOT analysis. If the issue is crucial, consulting an outside facilitator may be worthwhile.

Gather the Evidence

Often the best source of evidence is your own organization. You keep records of calls for service and your financial accounts. Those and other resources can give you valuable insights. Usually, internal data will provide a good baseline or a measure of the status quo.

Outside your organization, other sources of information are available. Professional and trade organizations are a good place to start. Suppliers will also give you information on comparative options and estimates of lifetime service costs. Do an online search. Despite all of the trash on the internet, there are also nuggets to be had. Learn how to use your favourite search engine to eliminate as much of the irrelevant material as possible. Do not be afraid to check organizations in outside jurisdictions.

Other excellent sources of information are libraries and your local college or university. Libraries have access to online databases that can search academic articles and other specialized material. Some of this can be intimidating to us if we are not used to using the facilities. Remember, a librarian can be your best friend. Contact your municipal librarian or visit a local college to seek expert advice.

Librarians can also help you navigate a wealth of statistical databases. Most provinces and provincial agencies collect

Remember, a librarian
can be your best friend.

and make available regional data. While most data are available to the public, some is limited to authorized agencies. If you work for a public service agency, it is likely that yours is one of those authorized agencies. The Statistics Canada website is also a valuable source of information.

Some colleges and universities have laboratories and research groups or institutes that focus on matters related to your office's mandate. Again, these can often be found through an internet search or by asking a local librarian for help.

Do keep in mind, however, that not all evidence is of equal value. Do not be afraid to be critical, or contrarian, especially if claims are at odds with your department's or your colleagues' experience. While not always the case, if something is too good to be true, it generally is. Ask yourself if the source is trustworthy. Is the agency presenting the data operating impartially or at arms-length, or does it have a self-serving agenda? Has the research or the publication gone through an external review process?

Organize the Evidence

Once you gather it, put your evidence together in an organized manner. Costing studies are easily presented in a spreadsheet. Other material can be presented in a table. Be sure to record the source of your information and keep track of where you found it. That way, if someone questions its veracity, you can refer them to the source.

A key element in presenting data is putting it in context. Remember, nothing means anything unless it is relationship to something else. Ask yourself, “compared to what?”

Is a three-minute average response time for calls for service adequate in your police or fire department? Can we drill down to priority calls to extract more precision? You can be assured that your supervisors, elected officials and others in the local community will ask.

A key element in presenting data is putting it into context.

Is a million dollars an appropriate price for an online registration system? Is it a necessary purchase or a colossal waste of funds if it is not implemented properly? Is our level of training adequate? Will training requirements change in the near future? If so, how?

These questions can only be answered by making reference to a comparable benchmark. What is the price range for goods and services in the marketplace? What are industry norms or standards for performance? Are there best practices against which you can compare your department or organization?

The more you engage in evidence-based decision making, the easier it will become. Knowledge is cumulative.

Review the Decision-making Process

Once you have done your analysis, it is good practice to review the entire decision-making process. What have you learned? How could the process be streamlined or made more efficient? The more you engage in evidence-based decision making, the easier it will become. Knowledge is cumulative. You will soon determine the best sources of information. You will discover how to make the process more efficient and how to minimize the likelihood of getting sidetracked.

While evidence-based decision making generally takes longer than other approaches, it has its benefits. Decisions based on hard evidence are more resilient in the face of scrutiny. We owe it to ourselves and the communities we serve to be more evidence-based in our thinking and application.

Taking a request to your boss or city council with strong external evidence is more likely to result in a positive decision. Presentations that show prior examples of success or that have reliable estimates of returns on investment are powerful. Finally, if someone challenges you, it is fair play to say that you have provided evidence to support your request. If they disagree, then ask them to show you their numbers.

What Others are Saying about *The Right Decision*

Municipalities are the engine of our economy and home to the majority of Canadians. The complexity of decisions and resource allocation in local government is growing rapidly. Preparing our staff with the skills to make evidence-based decisions which reflect the local context is essential. This readable and practical handbook is an excellent tool, accessible to staff at all levels, and a remarkable step in enhancing the performance of public servants at all levels of government.

Penny Ballem, MD FRCP, former City Manager, City of Vancouver; Clinical Professor of Medicine, University of BC

Although the focus of this manual is on evidence-based decision making, it also provides an important reminder that not all decisions are, or can be, based strictly on facts. Other factors need to be considered. We sometimes need to make the best decision, not the absolute correct decision, based on the situation, circumstances, internal and external factors, political environment, etc.

The processes outlined in *The Right Decision: Evidence-based Decision Making for Government Professionals* are a recipe for building a high performing team and creating a culture of continuous improvement, best practices and innovation.

Francis Cheung, P. Eng., Chief Administrative Officer, City of Langley

The Right Decision: Evidence-based Decision Making for Government Professionals is another in a series of works that are designed to ensure that governments provide services that are actually required, provide excellent value, and are delivered within an analytical framework. Staff in any government organization would benefit from the practical step-by-step approach to program development combined with the case studies from real-life projects.

George C. Duncan, Chief Administrative Officer, City of Richmond

My colleagues in the legal profession and I know that the best chance of winning a case depends on having the evidence to support our arguments. As a local government lawyer and a police board member, I have welcomed the growing emphasis in government spheres on making evidence-based decisions. *The Right Decision: Evidence-based Decision Making* should be required reading for every current and aspiring politician and their staff advisors at all levels of government in Canada—local, provincial and federal.

Lorena (Lori) Staples, Q.C., Lorena P.D. Staples Law Corporation and Saanich Police Board Member

Evidence-based decision making is becoming increasingly important as municipal Councils and staff wrestle with issues in a quickly changing, ever more complex world full of competing interests. This manual provides an excellent resource for those seeking to increase the role of evidence in municipal decision making.

David Stuart, Chief Administrative Officer, District of North Vancouver

This book is an excellent primer for evidence-based decision making. The language is clear; it is comprehensive and logical; and, there are plenty of examples to assist practitioners. Many new public servants would benefit from reading this book as they embark upon their careers.

Lori Wanamaker, FCPA, FCA, Deputy Minister, BC Ministry of Justice

If we could get more governments to implement evidence-based decisions, most operations would not have to contract out or outsource because no private company could compete.

Ken Wiesner, former Chief Administrative Officer in a number of municipalities and director of the Canadian Association of Municipal Administrators

Making the Right Decision

As a government service professional, you make crucial decisions every day that balance need with available resources. How should you approach these decisions, and how can you justify the decisions you make?

In this manual, Professor Paul Maxim, Fire Chief and Professor Len Garis, Professor Emeritus Darryl Plecas and legal analyst Mona Davies explore the what, why and how of evidence-based decision making.

What Others Are Saying About *The Right Decision*

Please see the inside back cover for full versions of these and other endorsements.

As government agencies become increasingly challenged to deliver programs with limited resources, the true test is finding the right balance between effectiveness and efficiency. This manual takes a comprehensive look at how success can be achieved on both fronts through evidence-based decision making and how it is the right tool to ensure the public gets the best value from our efforts and their tax dollars.

Vincent Lalonde, M. Sc., P. Eng, City Manager, City of Surrey

This readable and practical handbook is an excellent tool, accessible to staff at all levels, and a remarkable step in enhancing the performance of public servants at all levels of government.

Penny Ballem, MD FRCP, former City Manager, City of Vancouver; Clinical Professor of Medicine, University of BC

The processes outlined in *The Right Decision: Evidence-based Decision Making for Government Professionals* are a recipe for building a high performing team and creating a culture of continuous improvement, best practices and innovation.

Francis Cheung, P. Eng., Chief Administrative Officer, City of Langley

Staff in any government organization would benefit from the practical step-by-step approach to program development combined with the case studies from real-life projects.

George C. Duncan, Chief Administrative Officer, City of Richmond

This manual provides an excellent resource for those seeking to increase the role of evidence in municipal decision making.

David Stuart, Chief Administrative Officer, District of North Vancouver