issa | INTERNATIONAL SOCIAL SECURITY ASSOCIATION

# ISSA Social Security Research and Policy Manual

Supporting research in social security institutions in low- and middle-income countries

# ISSA Social Security Research and Policy Manual

Supporting research in social security institutions in low- and middle-income countries

International Social Security Association
Geneva, 2010

The International Social Security Association is the world's leading international organization bringing together national social security administrations and agencies. The ISSA provides information, research, expert advice and platforms for members to build and promote dynamic social security systems and policy worldwide.



The designations employed herein, which are in conformity with United Nations practice, do not imply the expression of any opinion on the part of the ISSA concerning the legal status of any country, area or territory or of its authorities, or concerning the delimitation of its frontiers.

While care has been taken in the preparation and reproduction of the data published herein, the ISSA declines liability for any inaccuracy, omission or other error in the data, and, in general, for any financial or other loss or damage in any way resulting from the use of this publication.

All modules of the *ISSA Social Security Research and Policy Manual* are available for download on: www.issa.int

International Social Security Association
4 route des Morillons
Case postale 1
CH – 1211 Geneva 22
T. +41 22 799 66 17
F. +41 22 799 85 09
E: issacomm@ilo.org
www.issa.int

ISBN 978-92-843-1191-0

# Preface

Around the world, social security administrations are continuously seeking improvements in their organizational and financial performance, with the objective of improving member services and benefits as well as of better extending social security to all. Innovations are needed to develop increasingly accessible and sustainable social protection systems which are socially inclusive and economically productive.

The worldwide membership of the International Social Security Association (ISSA) therefore promotes Dynamic Social Security approaches based on integrated, proactive and forward-looking policy choices. To be able to make such choices social security administrations need reliable evidence about the past and current performance of their own schemes, and projections about the future.

Social security administrations manage administrative registers that provide valuable information sources for statistical and actuarial analyses, especially if combined with other quantitative data from household budget surveys and other regularly-collected data managed by national statistical offices or government ministries. By joining forces with market research agencies, policy think-tanks and development agencies that use opinion polls and participatory assessments they can reach a better understanding of the drivers behind people's behaviour as consumers, voters and as actors in society.

Research activities require certain resources, but these should be regarded as an investment rather than as a cost by social security administrations, as they often generate excellent rates of return.

While the usefulness of social security research to policy-makers is often acknowledged in public debates, this is much less the case in social security administration practice, in particular in low- and middle-income countries. The ISSA has over the years repeatedly studied the reasons behind this situation.[1]

The explanations given include such justifications as the lack of appropriately trained researchers working in social security administrations, lack of resources to devote to research activities and lack of reliable data and information on which to base analytical studies. These reasons help to explain why more than half of the institutions that are members of the ISSA do not have an "in-house" research staff.

It should however be taken into consideration by the institutions concerned, that rigorous research is at the heart of evidence-based social security policy-making and scheme management. This implies that research can provide the evidence needed to help ensure that policies and schemes are:

- well-designed;
- appropriately targeted;
- likely to achieve their intended outcome;
- accurately costed; and
- reduce unexpected/unwanted outcomes.

Research can further provide social security policy-makers with information about what works, for whom, under what circumstances and how much it will cost – tomorrow and 30 years from now. It can also help to ensure that scarce resources are well-used and are not misdirected.

1. In particular, see ISSA (1998). *The role of research in social security.* Studies and Research, No. 25 (Geneva).

In order to respond to the various needs mentioned above, the *ISSA Social Security Research and Policy Manual* pursues three main objectives:

- To develop and strengthen the research capacity of social security institutions in low- and middle-income countries;
- To help, in turn, managers and directors in social security institutions of these countries to improve policy design, as well as implementation and administration of social security schemes, through setting research and policy activities in their institutions;
- To support research and analysis staff of member institutions, because by better understanding the social security stakeholders concerns, they will be able to focus their research on relevant policy issues and improve the quality and impact of their input.

The ISSA would like to thank the members of its Advisory Board for Policy and Research for its support for this project. It was the Advisory Board that initially recommended to the ISSA that the promotion of social security research among low- and middle-income countries be included as a priority in the ISSA's Policy and Research Programme due to its importance for the future of social security in these countries.

The modules of this Manual have been written by a range of experienced and qualified experts in the field of social security research. We would like to express our appreciation of the high quality of their work.

Finally, this Manual also aims to provide social security administrations and leaders, in particular ISSA members in low- and middle-income countries, with a practical manual that can motivate them to build adequate research resources and assist them in the design and implementation of successful research programmes.

In order to develop and improve this Manual further in its future editions, the ISSA Secretariat encourages feedback from readers and users – your comments and suggestions are most welcome!

**Hans-Horst Konkolewsky**
Secretary General

# Module objectives

**Module 1: Research and social security decision-making**
This module highlights how the impact of research will depend on the way it has been carried out and on the way its results are communicated. Social security organizations need to adapt constantly to realize improvements in organizational performance and programme outcomes. This adaptation may prove necessary in different circumstances, such as social security reform, policy implementation, administrative improvements and the extension of coverage. In this way, it is crucial to build a unifying framework for action to help social security administrations to better realize improvements in performance and desired programme outcomes.

**Module 2: The policy-research relationship**
This module expounds on how the policy-research relationship functions in social security institutions. Good information is at the heart of all good decision-making. Though the policy-research relationship is challenging, the benefits of a productive relationship more than repay the investment in developing it. Acknowledgement of the different roles and cultures of policy-maker and researcher – and the constraints and opportunities of each – is crucial for a constructive working relationship.

**Module 3: Management of the research process**
This module discusses the various stages in the management of the research process. An organization needs to know how it can adapt itself to make the whole process of procuring research and using research findings much simpler. This, in turn, will help to ensure that the research delivers useful outputs and that money is well-spent.

**Module 4: Statistical research on social security**
This module addresses statistical research on social security, including the indicators and data system, the various categories of data required as well the main collection and storage techniques. For achieving good governance, appropriate indicators and accurate, up-to-date and complete data are indispensable. Without them, the design, management, monitoring and reform of social security systems will not work properly.

**Module 5: Actuarial research into the future performance of social security**
This module focuses on actuarial research, encompassing the questions to be investigated, modelling, social budgets and various actuarial models. Quantitative actuarial research can provide plausible answers to many questions about future national social protection schemes as: How will future income and expenditures of the scheme develop? How does the scheme react to potential future demographic and economic developments? What are the financing and income replacement implications of proposed reforms?

**Module 6: Qualitative research**
This module outlines the basic concepts, practicalities as well as a variety of tools of qualitative research. Qualitative research is a uniquely useful method of finding out about people, what they think, feel, hope, believe and understand. It enables us to explore beneath the surface and to consider why people do what they do, think how they think and behave the way they behave.

**Module 7: Other forms of evidence – and how to use them?**
This module covers inclusive approaches to policy-making and other forms of evidence and the ways to use them. While quantitative sources of data are useful in producing some forms of evidence, qualitative sources play an important role, including helping to produce policies that work in practice. Inclusive and " joined up" approaches to policy-making can play an important role in helping policy-makers devise more appropriate policies and achieve more effective policy outcomes.

**Module 8: Evaluation research on social security**
This module describes the role of diverse types of evaluations in social security research. Good programme and policy evaluations assess the performance of programmes and policies, measure their impacts on individuals, families, communities and national development goals, and document the successes achieved, or the shortcomings. With evaluation information, policy-makers and programme level decision-makers are able to direct limited resources to where they are most needed and most effective for their contributing members and the wider communities.

# Table of Contents

# Boxes and Figures

## Module 6



## Module 8

# Authors

**Catherine Bochel** is Principal Lecturer in Social Policy at the University of Lincoln, United Kingdom. (Module 7: Other forms of evidence – and how to use them?)

**Florence Bonnet** is a Statistician in the Social Security Department of the International Labour Organization. (Module 4: Statistical research on social security)

**Theo Butare** is Senior Social Security Technical Specialist at the International Labour Organization, based in Yaounde, Cameroon. Formerly, he was a Senior Programme Manager at the ISSA in charge of the Statistical and Actuarial Programme. (Module 4: Statistical research on social security)

**Sue Duncan** is an independent consultant with over thirty years' experience working in the public sector, at the centre of Government in the Cabinet Office and Treasury, United Kingdom, and in social policy departments. (Module 2: The policy-research relationship; Module 3: Management of the research process)

**Wouter van Ginneken** is a consultant working with the ISSA. He was a Senior Economist at the International Labour Organization's Social Security Department. (Module 1: Research and social security decision-making)

**Aidi Hu** is a Social Security Specialist at the Social Security Department of the International Labour Organization in Geneva in charge of Asian countries, with a particular focus on China. (Module 4: Statistical research on social security)

**Warren McGillivray** is a Policy Associate with the Caledon Institute, Ottawa, Canada. He was Chief of the Studies and Operations Branch of the ISSA from 1993 to 2004. (Module 5: Actuarial research into the future performance of social security)

**Roddy McKinnon** is Editor of the *International Social Security Review* and Publications Manager at the ISSA. (Module 1: Research and social security decision-making)

**Timo Voipio** is Senior Adviser for Global Social Policy at the Ministry for Foreign Affairs of Finland. Formerly, he was a Project Manager in the ISSA's Policy and Research programme. (Module 6: Qualitative research)

**Robert Walker** joined the University of Oxford, United Kingdom, as Professor of Social Policy in April 2006 when he also became a Fellow of what is now Green Templeton College. (Module 8: Evaluation research on social security)

# Acknowledgments

This Manual was produced by the ISSA's Policy and Research programme, under the coordination of Roland Sigg.

Timo Voipo was responsible for the project management and the editing of the Manual, and Pierre-Alain Roch provided editorial support for the final version.

MODULE 1

# Research and social security decision-making

RODDY MCKINNON – WOUTER VAN GINNEKEN

## 1.1 Matching the research needs, outputs and impact

The key purpose of social security research is to help social security policy-makers and scheme managers perform their work more smoothly and with better results. Thus, this Manual seeks to make life easier for both managers and researchers of social security.

Rigorous research is at the heart of evidence-based social security policy-making and scheme management. Research can provide evidence to help ensure that policies and schemes are:

- well designed;
- appropriately targeted;
- likely to achieve their intended outcome;
- accurately costed, and;
- reduce unexpected/unwanted outcomes.

The intention of research is that it should provide social security policy-makers with information about what works, for whom, under what circumstances and how much it will cost. It helps to ensure that scarce resources are well used and are not misdirected. Research can help inform decisions about different policy options to ensure that those which are selected are best suited to the particular situation.

The ISSA conducted a survey in 2006 among its over 350 member organizations. The prominent concerns for most members were:[1]

- the financial implications of *demographic ageing* for social security;
- implications of *globalization* for national social security programmes;
- *extension* and *improvement of access* to social security; and
- improvement of the *organizational performance* of social security administrations.

Social security institutions need to have access to up-to-date knowledge and relevant research evidence to respond to these challenges. But also other stakeholders – such as governments, social partners, civil society organizations and the social security clients themselves – need access to social security research.

This Manual gives leads and guidelines to all these social security "stakeholders" as to how they can obtain relevant knowledge to make good decisions and how to evaluate delivery and implementation. This Manual will also be very useful to those who undertake research, because by better understanding the stakeholders' concerns, they will be able to focus their research on relevant policy issues and improve the quality and impact of their output.

### 1.1.1 The ISSA Social Security Research and Policy Manual

This Manual is, first of all, aimed at managers and directors in social security institutions. They need quality evidence and knowledge to support decision-making. But their decisions are taken in a political context in which governments as well as other stakeholders, such as social partners, client groups and international actors, play an important role.

Managers and directors in social security institutions are always looking for ways to improve policy design, as well as implementation and administration of social security schemes. Changing policy design to extend and improve social security coverage is one of the main concerns in low- and middle-income countries (LICs and MICs). It is in this context that the ISSA Manual seeks to make a contribution.

1. ISSA (2006), with some regional variance.

The extension of *legal* coverage is not sufficient to reduce the everyday *social* insecurities of real people. It is also important to know what is the real degree of law enforcement. As some factors that cause social insecurity always fall outside the scope of legal regulation, much is left to the discretion of the front-line staff ("street-level bureaucrats"). Policy changes need more than just changing the legislation.

A key issue here is serious analysis of whether the existing benefits provided correspond to the priority needs and the contributory capacity of the working population and their families – in particular, those working in the informal economy. For instance, should the emphasis be placed on providing long-term benefits, such as old-age pensions, or on meeting more immediate and productive needs for health care or – in some cases – housing? (Baruti, 2008)

In turn, blockages created by a lack of flexibility in the existing social security schemes, while often linked to real pragmatic administrative and cost considerations, can prolong the situation in which social security remains a privilege accessible to only a few – and sometimes not even to all those who are eligible.

There is a genuine interest in improving the cost-effectiveness of social security administrations and in using performance management tools, such as activity-based costing, the balanced score card and benchmarking. New research and the interpretation of existing research outcomes are needed to monitor, assess and implement such management tools that can help realize the expected performance goals.

This Manual is also aimed at researchers both within social security institutions and outside, in the academic community and in research institutions. For research to have an impact on policy, two things are needed (ODI, 2005):

- *Relevance* – both substantive focus and operational usefulness. Topical and operational relevance can be enhanced by including clients and local social security staff members in the research process. Including policy-makers in the research process can help ensure topical and operational relevance for public policy.
- *Credibility* – not only with regard to policy-makers, but also with the media and the general public. Research outcomes are more credible when they are based on solid evidence, rigorous methodology and – where appropriate – pilots that have worked in practice. The policy impact of research is also dependent on the way the results are communicated: short, clear and jargon-free documents and demonstration through real-life case studies generally work best.

### 1.1.2 Dynamic Social Security (DSS) and the demand for research

Social security organizations need to adapt constantly to realize improvements in organizational performance and programme outcomes. The growing pressure for rapid and large-scale adaptation is brought about by the new challenges of ageing societies, globalizing labour markets and their implications for employment, poverty and migration.

This adaptation may prove necessary in different circumstances, such as social security reform, policy implementation, administrative improvements and the extension of coverage. McKinnon (McKinnon, 2007) proposes a unifying framework for action to help social security administrations to better realize improvements in performance and desired programme outcomes. Research can play an important role in realizing the goals of Dynamic Social Security (see Box 1.1).

### 1.1.3 The organization and impact of research

A great variety of statistics and research reports is available in every society. Many of these are relevant to social security administrations. However, social security administrations also need specific research outputs that they can either commission to outside researchers or undertake within their own organization. The impact of such research will depend on the way it has been carried out and on the way its results are communicated.

The development and regular collection of good social security data are the first requirements for relevant and credible research outputs.

Statistics are good when they are based on relevant statistical concepts and on reliable data sources. They are all the more useful when they are comparable over time and, if possible, between countries.

The ILO (ILO, 2005) is developing a statistical knowledge base that includes three types of data:

- *Social Expenditure and Performance Reviews* (SEPRs), providing detailed information on the (mainly financial) performance of national social protection schemes;
- *Scheme-Specific Performance Indicators* (SSPIs), subdivided into legal, governance and financial indicators;
- *Social Security Inquiry* (SSI) that collects internationally comparable social security statistics. One of the main challenges is to develop and collect new statistics on social security coverage (Van Ginneken, 2007).

Social security administrations use available research outputs for their decision-making and for the improvement of their performance. Many of those outputs are focused on "good practice" experiences and are often produced by international organizations such as the ISSA, the ILO, the World Bank and regional banks. They may also be produced by national or regional research institutes with a specific focus on social security or where social security is one of the issues investigated. These research results will probably be digested by the research units of social security institutions, but they are often not enough for specific decision-making and improved performance.

Faced with specific policy issues, social security administrations will therefore have the choice to (a) subcontract research to outsiders, or (b) undertake it within their own organization.

Some minimum research capacity should be available in social security administrations so as to successfully subcontract research to outsiders, to prepare policy decisions and to properly interpret and assess the findings from commissioned research.

The advantages of having in-house research staff with up-to-date skills and capacity are that requests for policy advice can be responded to quickly and the implications of policy issues will be clearly understood.

Possible disadvantages are that it may be difficult for in-house research personnel to keep themselves up to date with the newest research methods and knowledge in the academic world and, with time, it may also become difficult for in-house researchers to think "outside the box".[2]

In practice, the solution would often be to have at least a small research staff in-house and to have a budget that is also available to commission outside research.

2. …and encouraged, or even allowed, to do so.

### Box 1.1. Dynamic Social Security (DSS)

The ISSA uses DSS as a unifying conceptual framework to guide:

(1) *action* in social security policy towards realizing high-performing social security institutions and universal access to sustainable benefits and services;

(2) *research* towards identifying policy solutions that are forward-looking and proactive but also sensitive to national conditions.

**What are the priorities of DSS?**

Priorities are to:

(1) *defend* existing models of social security provision, but also to realize measurable ***improvements*** in social security policy outcomes;

(2) encourage knowledge transfer on administrative and policy *good practice*;

(3) promote wider institutional *partnerships* among social policy agents;

(4) create *open-mindedness* to the possibility of new or complementary programmes.

**How can research contribute to DSS?**

By identifying good practice solutions and feasible policy-reform pathways towards:

(1) realizing *integrated* and *coherent* social security systems;

(2) creating more *inclusive* societies and more *productive* economies;

(3) enlightening policy-makers on accessibility:
- relevance of benefits
- proximity of services
- client satisfaction
- value for money

(4) finding ways to improve the financial, political and social sustainability of social security administrations and the programmes they deliver;

(5) documenting evidence of the social and economic benefits of social security, e.g. in the absence of social security, many people would not have access to health and education services and they would not benefit from programmes that prevent and protect against risks, and which help them rehabilitate as well as reintegrate into society, the economy and the labour market.

DSS also highlights three basic characteristics of policy pathways towards desirable social security outcomes. Social security should be (a) proactive; (b) innovative; (c) credible.

Research can help to identify proactive measures that encourage prevention and mitigation, for example through promoting healthier lifestyles as well as rehabilitation and skill improvement. Research findings can also illustrate how *innovative linkages*, for example between statutory social security and community-based social protection mechanisms, could offer a promising new approach to extending health-care coverage. Policies are *credible* when they are founded on relevant and quality research and when they are sustained by solid economic, political and social support.

The management of the research process is most important and will be discussed in various parts of this Manual. Good research management starts with the drafting of the *terms of reference* (TOR), which indicate the expected research outputs, the time line and the budget available. This is a process that can only be carried out within the social security organization.

- WITHIN: If the research is to be carried out *within* the organization:
  - the budget implications for the organization will have to be assessed;
  - the time schedule will have to be established;
  - the person(s) involved will have to be selected (motivation and training play an important role here).

- OUTSIDE: If the research is to be carried out *outside* the organization:
  - it is crucial to make the right choice of researcher or research institute;
  - part of this choice will be determined by the fee that is proposed, but scouting around for the right person is a most worthwhile activity.

Research management also requires constant monitoring during the research process: by regular encouragement and by making sure that comments on intermediate outputs are provided promptly and that the final product is properly disseminated and used.

As noted earlier, the impact of research can be enhanced by taking a participatory approach, i.e. by including clients and policy-makers in the research process.

The policy impact of research is also dependent on the way the results are communicated:
- A first requirement is that the main research outputs and policy conclusions are summarized in clear and jargon-free language.
- Secondly, the impact on policy-makers can be enhanced through discussions and short presentations.
- The results should be communicated to the media through press briefings and interviews, so that the results are known to interested stakeholders and the general public.
- Finally, the full study should be quickly available in printed form and on the website.

See Box 1.2. for an example of the impact of a research strategy in Nicaragua.

## 1.2 Dynamic Social Security

### 1.2.1 Fostering an innovative vision of social security

In recent years, the ISSA has been confronted with a conceptual challenge. In undertaking the routine observation, analysis, and communication of global developments and trends, it was clear that the ISSA's conceptual tool kit was no longer sufficient to meet this task in full: the pace and scope of change under way across social security policy and practice worldwide – not least in relation to demographic change and the globalization of markets – was challenging the ISSA's ability to confidently identify, cluster and promote "good practice" developments and trends in social security.

This identified conceptual challenge is largely rooted in the day-to-day policy challenges and questions faced by social security administrations. What must social security administrations do to better extend coverage to all? How can the expressed, but often different, needs of all citizens and workers be better met by social security? In turn, what are the operational needs of social security administrations and their stakeholders? And how must administrations adjust to better satisfy all of these? Where do the management and policy solutions lie?

To address these questions, the ISSA is fostering an innovative conceptual framework for social security: *Dynamic Social Security* (ISSA, 2007; McKinnon, 2007, 2009).

### Box 1.2. Nicaragua pension reform repealed based on research[3]

In the early 2000s, the World Bank and the International Monetary Fund (IMF) advised Nicaragua to privatize public services. As part of this process, in 2002 Nicaragua passed a law aimed at transforming the pay-as-you-go (PAYG) pension scheme into a new scheme with individual savings accounts, to be managed by private pension fund administrators.

However, as a result of intensive research and debate facilitated by the Nicaraguan Social Security Institute (INSS), this legislation was repealed four years later, in 2006.

The research strategy the INSS used was to first create a High-Level Commission with representation of the Ministry of Finance and Public Credit, the Central Bank and the INSS, together with expert advice of the ILO. The Commission partners then linked their ITC systems so that the projection models could make use of the statistical data of the INSS as well as the economic and financial data of the Central Bank and the Ministry of Finance.

Using actuarial valuations and financial/economic projections, the researchers demonstrated that the reform would create an actuarial deficit of NIO 77 billion and an unsustainable level of public indebtedness (mainly because of the financing of minimum benefits), which as a whole would have jeopardized the macroeconomic stability of the country. Once the research results had been communicated and conclusions agreed with social partners, the Parliament of Nicaragua had no other option than to repeal the law.

The INSS has continued to undertake technical studies including actuarial valuations with a view to assess the sustainability of the PAYG pension scheme. The studies have shown that the scheme is sound and that its reserves have grown considerably as a consequence of the extension of coverage, efficient management and the establishment of investment policies. The ongoing research has also helped the INSS to identify potential financial problems by 2020.

**Conclusion**

Legal instruments must always be supported by financial and economic analyses and be based on a social consensus. This is even more evident when laws have an impact on social issues, as in the case of pensions. Social security organizations must play a decisive role in evaluating the impact of reform measures. In doing so, research activities and technical work should never be neglected and should be part of their regular activities.

## 1.2.2 How do we define Dynamic Social Security?

Dynamic Social Security is defined as:

*Policies and processes geared to better ensure accessible and sustainable social protection systems that not only provide protection, encourage prevention, and support rehabilitation and reintegration but also contribute to better realizing socially inclusive and economically productive societies.*

Key characteristics of Dynamic Social Security are:

- Accessible
- Sustainable
- Proactive
- Innovative
- Socially inclusive
- Economically productive

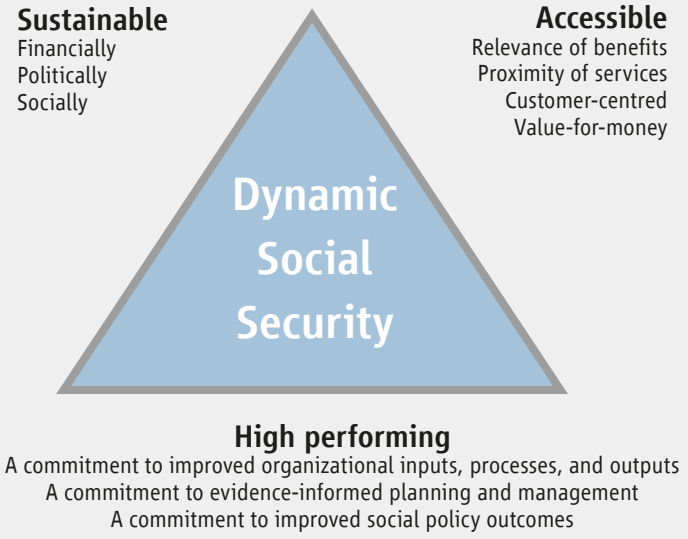
3. Information provided by Dr Roberto López Gómez, Executive President, INSS; and the ISSA Secretariat.

Importantly, and far from being prescriptive, Dynamic Social Security presents a pragmatic policy vision. In the pursuit of Dynamic Social Security responses, it is for the responsible ministries and government agencies, as well as social security administrations, to define the policy content and organizational processes appropriate to their own national or sectoral circumstances, mandates and requirements.

### 1.2.3 What are the aims of Dynamic Social Security?

Dynamic Social Security aims to support existing social security administrations in their ongoing endeavours to improve organizational inputs, processes and outputs. It also promotes efforts to realize wider access to social security, with an emphasis on improving social security policy outcomes. In turn, it seeks to play a role in better ensuring the financial, political and social sustainability of social security administrations and the programmes they deliver (see Figure 1.1.).

Figure 1.1. The core aims of Dynamic Social Security




The immediate aim of Dynamic Social Security is to support the development of high-performing and policy outcome-oriented social security administrations. In the longer term, the objective of Dynamic Social Security is to advance, at the very least, basic income security and access to health care for all.

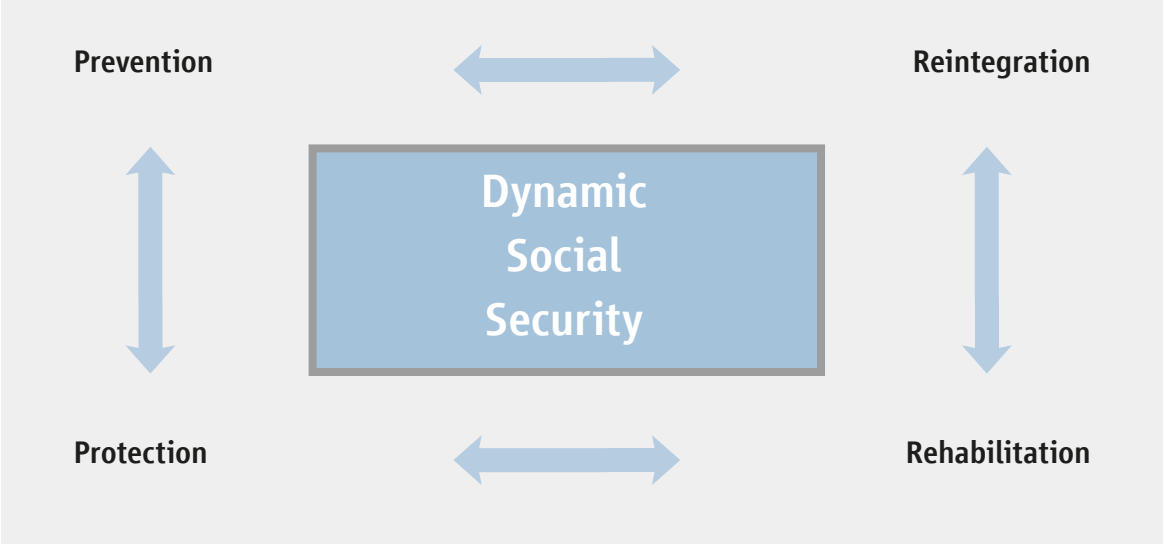
### 1.2.4 What does Dynamic Social Security imply for social security administrations?

A key message of Dynamic Social Security is that a change in mindset is required. It is essential that all social security administrations see themselves as proactive agents in the bigger socio-economic picture.

This proactive stance may take the form, for example, of policy pathways (Figure 1.2.) and measures that encourage prevention through heightening individual awareness about risk behaviours or that inculcate attitudes that give increased priority to healthier lifestyles. Proactive approaches can also mean promoting rehabilitation and skills promotion for those once deemed incapable of any work, or prioritizing the reinsertion into formal work of those marginal to the labour market.

Of course, this labour market-oriented outlook is not applicable in all instances: for some individuals, a return to work is not an option, while for others it may never be an option. Furthermore, it cannot be ignored that work in the informal economy accounts for the lion's share of employment in many countries.

Figure 1.2. Policy pathways to Dynamic Social Security




As the challenge of high levels of informal employment underlines, realizing improved social security outcomes is dependent upon the coherent and integrated actions of many actors. Yet building such coherent and integrated solutions, even within the limited scope of public institutions and government ministries, is always testing. A further challenge presented by Dynamic Social Security, therefore, is one of improving knowledge about how to better include all legitimate actors from civil society, the workplace and private institutions.

### 1.2.5 Dynamic Social Security: Linkages, bridges and frontiers

If coverage is to be improved, social security administrations must define their role in relation to, and work in partnership with, other social actors. By fostering innovative policy design linkages among programmes that are complementary to one another, and in ways that seek to avoid perverse crowding-out effects, it should be possible to reduce social insecurity. It should also be possible to improve the robustness of social protection programmes, including – not least – civil society initiatives.

Research findings (Coheur et al., 2007) illustrate how innovative linkages between statutory social security schemes and community-based social protection mechanisms offer a promising new approach to extending health-care coverage. Through building innovative administrative bridges between civil society initiatives and statutory social security, it should also be possible to allow movement from voluntary to rights-based coverage.

In practice, such a bridge could provide the registered members of small occupational or workplace welfare funds with subsidized access to a more extensive basket of benefits and services under statutory national social security programmes. Nonetheless, this raises an important issue. How should we define the frontiers of action of conventional social security administrations?

Although some administrations' mandates are rooted in income replacement to mitigate the impact of an insured risk event, the roles of other social security administrations lean more clearly in the direction of social development and poverty alleviation. Others still, like provident funds, place an emphasis on individual savings and family responsibility. Different models of social security, like differences in philosophy and mandate, are a reality and due consideration must be given to the practical implications of these.

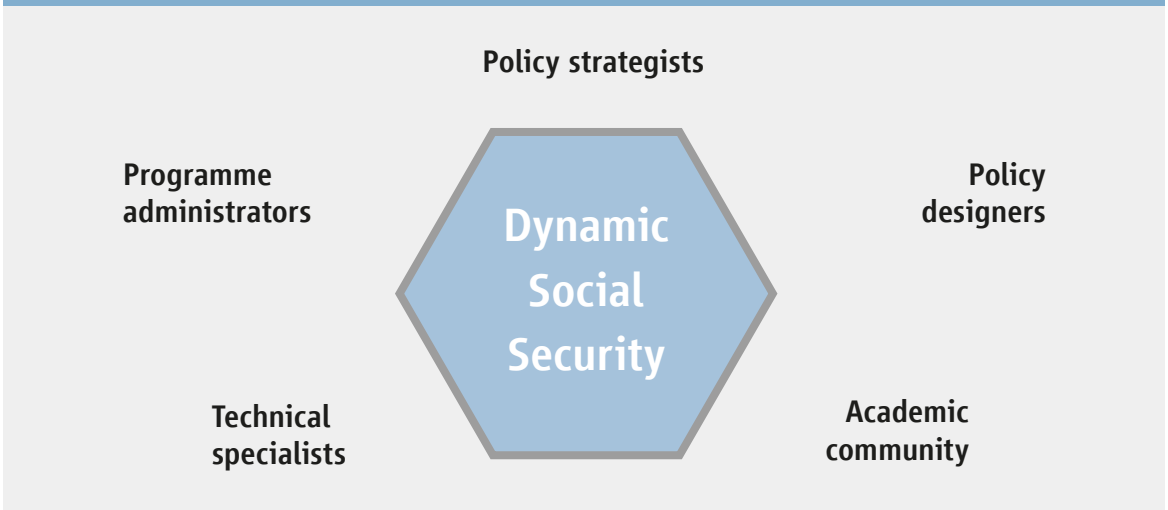
Therefore, social security's frontiers will remain, at the national and local level, defined by the roles performed by existing national and local social security programmes, and by their respective mission statements. Notwithstanding this, the possibility of these frontiers continuing to evolve must be looked upon as both likely and necessary.

### 1.2.6 Three reasons why Dynamic Social Security is important

Dynamic Social Security offers a common research lens (Figure 1.3.) through which social security administrations may observe international social security developments and trends. This common lens may also be shared by actors in the wider policy design and research field.

Dynamic Social Security provides a shared tool by means of which to influence attitudes about the direction and pace of change in social security policy and practice. In broad terms, the aim is to influence the path of change towards realizing improvements in administrative performance and social security policy outcomes that also have the capacity to impact positively on the wider needs of economies and society.

Figure 1.3. Dynamic Social Security as a common lens




Dynamic Social Security articulates a positive message of proactive engagement and win-win opportunities. This message – supported by emerging research evidence – may contribute to help mediate social and political tensions. The message to underline is that the extension of social security coverage is not only good for the advancement of social justice and to the realization of more inclusive societies. It is also good for national economy growth, not least by helping to better realize the productivity of those previously uncovered and through enhancing aggregate domestic demand. In a context where social security programmes in developing countries are often criticized for not reaching out enough to uncovered populations, this is vital.

### 1.2.7 Promoting policy research and diffusion

Dynamic Social Security can be understood as a means to better research and improve diffusion of social security "good practices". This is important because it should help better ensure the sustainability of social security administrations; and the identification, diffusion, risks and implementation of "good practice" improvements should help reinforce the relevance, effectiveness and credibility of benefit programmes as well.

Indeed, important political, administrative, and financial challenges confront many social security administrations. As an upshot, the reform process is often drawn-out and difficult. Notwithstanding all of this, the objective of Dynamic Social Security, as a conceptual framework, is to influence attitudes regarding the direction of reform in policies and programmes in order to better realize improvements in social security organizational inputs, processes and outputs, as well as the desired social security programme outcomes.

Consequently, as a research-oriented approach geared to help ensure more socially inclusive and economically productive societies, the potential of Dynamic Social Security to help better realize improvements in social security in lower- and middle-income countries cannot be ignored.

## REFERENCES

Baruti, E.E.M. 2008. *Social security reforms in Tanzania: The role of social security institutions in the fight against poverty* (Conference paper, 33rd Global Conference of ICSW, Tours, 30 June–4 July). Utrecht, International Council for Social Welfare.

Coheur, A. et al. 2007. *Linkage between statutory social security schemes and community-based social protection mechanisms: A promising new approach* (Technical report, No. 9). Geneva, International Social Security Association. <http://www.issa.int/content/download/40633/790295/version/9/file/TR-09-2.pdf> (accessed on 29.10.2010).

Ginneken, W. Van. 2007. "Extending social security coverage: Concepts, global trends and policy issues", in *International Social Security Review*, Vol. 60, No. 2–3.

ILO. 2005. *ILO social security inquiry: Manual.* Geneva, International Labour Office – Social Security Department.

––. 2006. *Social security for all: Investing in global social and economic development. A consultation* (Issues in social protection discussion paper, No. 16). Geneva, International Labour Office – Social Security Department.

ISSA. 2006. *In support of new ISSA: Survey of needs and interests of member organizations* (Draft Report). Geneva, International Social Security Association.

––. 2007. *Developments and trends: Supporting dynamic social security.* Geneva, International Social Security Association. <http://www.issa.int/content/download/39150/763035/file/2DT07.pdf> (accessed on 29.10.2010).

McKinnon, R. 2007. "Dynamic social security: A framework for directing change and extending coverage", in *International Social Security Review*, Vol. 60, No. 2–3.

––. 2009. "The ISSA and dynamic social security: Global social governance and local social action", in *Global Social Policy Journal*, Vol. 9, No. 1.

ODI. 2005. *Lessons so far. Evidence: credibility and communication.* London, Overseas Development Institute – Research and Policy in Development. <http://www.odi.org.uk/RAPID/Tools/Lessons_Evidence.html> (accessed on 29.10.2010).

MODULE 2

# The policy-research relationship

SUE DUNCAN

## 2.1 The big picture: Why is research important?

Good quality research and analysis can help to ensure that policies are properly costed and targeted and that they deliver what is intended. It can also provide feedback on how policy and delivery are working, so that mistakes can be identified early and rectified.

It is not only conventional research that is useful: wider evidence, including the views of consumers, practitioners, academic experts and those who actually deliver policy, can all help to ensure that policy is soundly based. The challenge is to draw effectively on these different sources.

The risks of *not* using research and other evidence are great: policy failure can be costly – both financially and socially. It also carries a reputational risk for the organization concerned. Often, policy decisions will have a political underpinning, but transparent decision-making where the supporting evidence is clear is an important element of the democratic process. It encourages wider public debate which can inform organizational decision-making and encourage new ways of thinking about social security policy.

### 2.1.1 For whom is social security research?

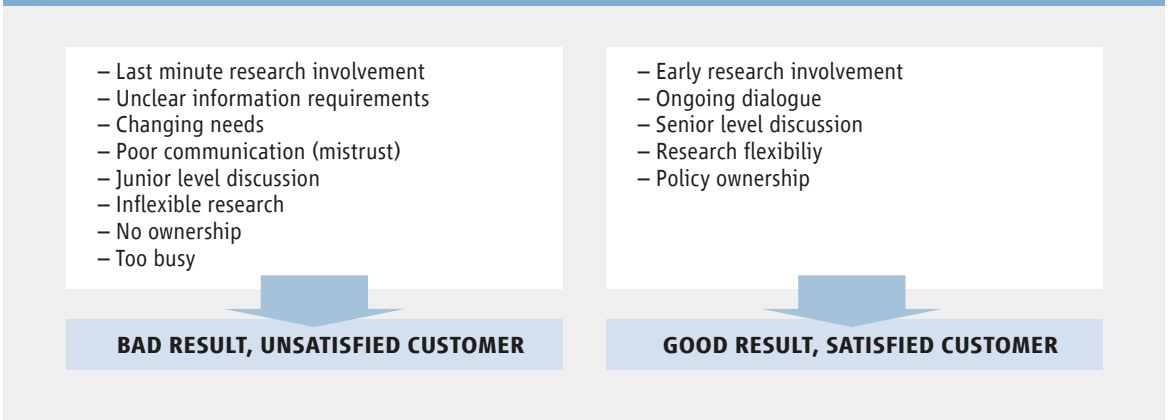
Social security decision-makers are important stakeholders for social security research, but its audience goes much wider. Crucially, it can inform wider policy debate and can help to ensure that social security organizations are accountable to consumers. Other important stakeholders are practitioners, interest groups and those responsible for delivering policy. All these people can potentially make use of research to do their jobs more effectively.

### 2.1.2 How does the policy-research relationship work in social security institutions?

Good information is at the heart of all good decision-making. Though the policy-research relationship is challenging, the benefits of a productive relationship more than repay the investment in developing it (see Figure 2.1.).

The relationship works well when it is close and when decisions are iterative, involving both parties at every stage. Researchers are not just passive purveyors of "facts", but they can use their research and analytical skills in contributing to the creative process of policy development and review. This said, acknowledgement of the different roles and cultures of policy-maker and researcher – and the constraints and opportunities of each – is crucial for a constructive working relationship.

Figure 2.1. The policy-research relationship



### 2.1.3 Conflicting timetables

Policy timetables are often constrained, while good research can rarely be carried out very quickly. If research is going to be useful, busy policy-makers need to take time to engage in the process. Without their input, research may not meet their needs.

Proper forward planning will help to ensure that research is available when it is needed. Similarly, research needs to be designed with users in mind. This can mean designing research so it delivers interim results or making trade-offs between scope and timing. Decisions are made on the best available evidence and cannot always wait for the fully comprehensive answer.

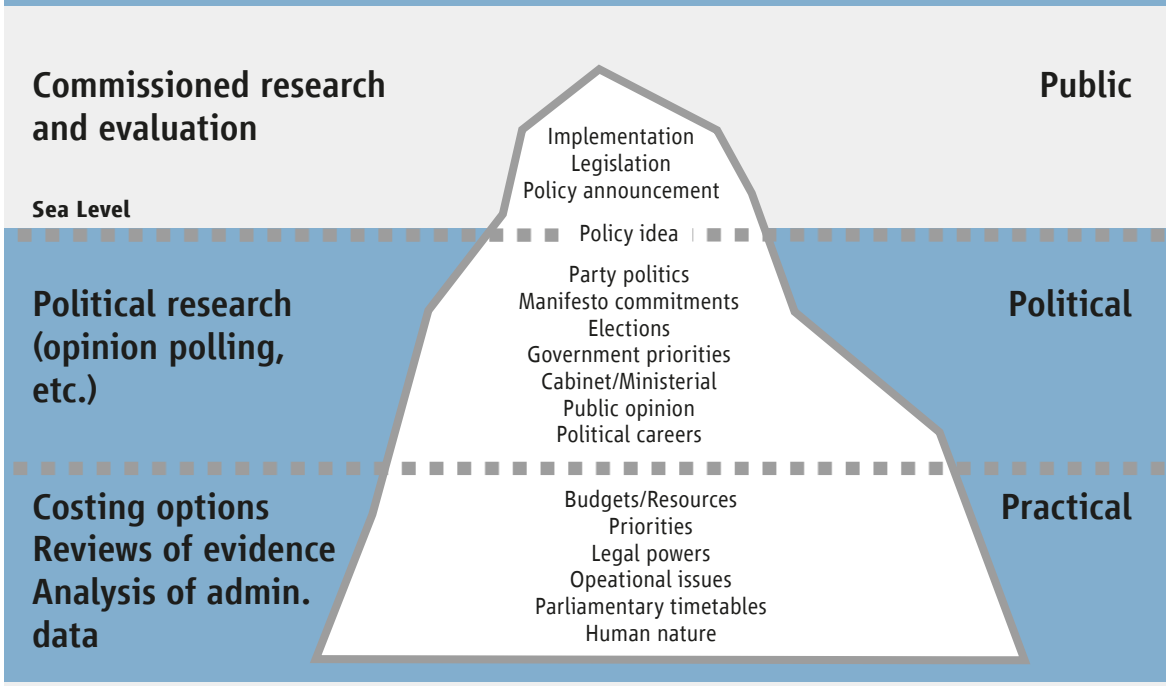
### 2.1.4 The challenges of using social science to inform policy

Social science research rarely generates the wholly conclusive answers that policy-makers ideally seek. This reflects the nature of the social world, but can be frustrating for decision-makers. In addition, knowledge is of course never complete and absolute: there are often gaps and uncertainties in our knowledge base where research has not been conducted or produces apparently conflicting findings, and new research brings new insights which can challenge conventional thinking. These limitations have to be recognized if the insights which research can bring are to be gained.

### 2.1.5 Influences on the policy research agenda

In the business of making policy, much activity goes on "below the waterline", often unseen by the public. Political priorities, resources and external influences, such as pressure groups and the media, can all affect the balance of power in the policy-research relationship and the influence research can have on the decision-making process. This can be frustrating for researchers, because it can constrain the influence of research, but it is a fact of life and reflects the dynamic nature of the policy process. Section 2.2 discusses this in more detail. Figure 2.2 below illustrates the complexity of the processes involved in decision-making in social security institutions.

Figure 2.2. The policy iceberg



## 2.2 Influences on the policy research agenda

Policy research and evaluation in social security administrations, as well as in government generally, tend to reflect current policy concerns. This is partly a value-for-money issue – organizations fund the research they need for doing their job – but it can also be a reflection of the tendency in politics to focus on the short term. This tendency towards short-termism can lead to important information gaps when research, which is not an immediate priority, is not commissioned.

Policy problems not on the current policy agenda do not go away and ways need to be found to anticipate and address future information needs. It may not be possible to collect data quickly enough when the problem re-emerges as a priority or when it can no longer be ignored.

Social security administrations have to acknowledge the existence of a problem or issue in order to respond. This can sometimes be politically difficult. Research can aid in this process, but the initial impetus may come from non-public-sector-funded research or from non-research sources. Publicly funded research may not be initiated until later (hence the importance of the wider research community).

Political parties have their own agendas and priorities, and this can affect research funding in social security institutions. Issues not on the current political agenda tend to get little public funding but may be picked up by other funders, e.g. charities, pressure/interest groups and universities.

Policy problems or policy failures may be highlighted by the media. Media coverage can bring issues higher up on the policy agenda and/or highlight information gaps and hence create a demand for research.

Pressure/interest groups working to represent and highlight the concerns of particular groups may be effective at grabbing media and political attention. They sometimes conduct their own research or draw on case studies to make a point. They can lead a social security institution or commission to conduct its own research to clarify the nature of the problem.

Another important actor in social security policy-making in developing countries is the foreign-aid donor community. It may be that social security provision, at least in the short term, is funded fully or partially by a donor who will significantly influence the design of the policies.

### 2.2.1 Linking in with the wider research community

Social security administrations should draw on wider expertise to inform policy. Research and thinking generated in universities and research institutes and research funded by other organizations are often a valuable source of data. They can provide new perspectives and help to share the burden of research funding. Time invested in developing wider research networks will enable access to a wide range of research and thinking, which can fundamentally improve the quality of decision-making.

## 2.3 Establishing an evidence-based culture

### 2.3.1 People

There needs to be a willingness and ability to use research in the organization. For this to happen, skilled social researchers who understand the policy process and policy-makers who understand how to use research and analysis are needed. The close working relationships that are essential to an evidence-based culture depend on an understanding of the two roles and how they work together.

Training is important in equipping staff in new ways of working, and the necessary culture change will need organizational commitment with leadership from the top and buy-in at the most senior level.

Those who have limited experience of using research may be sceptical of its value, concerned about loss of control or fearful that research may bring bad news. Such concerns need to be addressed if people are to recognize the value of research.

### 2.3.2 Resources

Suitably qualified staff within the organization will need to be supported by external research resources. It is rarely cost-effective for organizations to maintain an in-house fieldwork capability, and it is often more efficient to contract out some elements of the research function.

A budget will need to be allocated to enable commissioning of external research and for purchasing resources needed by in-house researchers (e.g. software packages for conducting data analysis).

Some resources will also need to be committed to a knowledge management system: without a means of storing information in an accessible form, investment in data collection may be wasted. It is essential that the system for managing knowledge reflects organizational work processes or it will not be fully used.

### 2.3.3 Organizational processes

For an evidence-based culture to thrive, the organization must encourage and, in some cases, require the use of research information. Research must be embedded in organizational processes so, for example:

- the identification of research priorities must be an integral part of annual planning cycles or there is a risk it will be overlooked;
- there need to be processes to ensure that decision-making is informed by evidence. This can be achieved by requiring all policy documents to cite the evidence which has informed the policy proposal or decision;
- administrative data must be routinely collected and collated. At the same time, it needs to be ensured that:
  - those responsible for administering the benefit understand the importance of reliable data and that they are also aware of the uses to which the data are put;
  - those responsible for the policy area that the data inform are charged with stewardship of the data collection process.

Unless organizational processes encourage the use of evidence, there is a risk that research and analysis will be seen as additional burdens rather than as aids to effective decision-making.

### 2.3.4 Strong policy-research relationships

If research is to be both useful and used, a strong policy-research relationship is essential. There needs to be a partnership which:

- recognizes the distinctive roles of each, but values the contribution of both;
- understands what each can contribute to the process of policy development, review and implementation;
- recognizes the importance of regular communication, consultation and information sharing;
- takes joint responsibility for an evidence-informed approach;
- sees policy as a collaborative and iterative process, where both researcher and policy-maker contribute to the shaping of research and the formulation of policy at key stages.

### 2.3.5 Research utilization

The means by which research findings enter the policy chain are complex, and a straightforward relationship between a research insight and a policy change is rare. More often, research findings influence policy incrementally over time; therefore, in many cases research is best viewed as a medium- to long-term investment.

The following are some examples of how research enters the policy chain:

- it can feed directly into decision-making for policy or practice;
- it can lead to changes in knowledge, understanding or attitude which can stimulate new insights into problems or suggest new policy solutions;
- it can influence large-scale shifts in thinking, leading to major changes in the way problems are conceptualized or the approaches to tackling them;
- it can be used to demonstrate that change is needed or to create the political climate in which change can take place;
- it can be used to legitimize particular courses of action;
- often, research and policy feed into each other as part of an iterative process; close policy-research relationships tend to facilitate this approach and, though very effective, the research impact can be difficult to detect; or
- it can address a previously identified problem and generate solutions.[1]

## 2.4 How can research help?

Research can inform policy and delivery in a number of different ways.

### 2.4.1 Understanding the big picture

Policy does not take place in a vacuum: Policy design and implementation need to take account of the wider economic and social context in which they are set and to anticipate and respond to economic and social change. Keeping track of social and economic trends allows social security institutions to monitor change in the size or nature of social problems and identify emerging issues and social problems. Monitoring trend data minimizes the risk of unanticipated policy problems and provides an early warning of emerging issues that will need a policy solution.

It is important to keep track of trends such as:

- changes in population (e.g. overall numbers);
- age structure (e.g. is the population ageing or growing younger?);
- migration, immigration, emigration and regional population balance;
- broad social changes (e.g. household formation; types of household; numbers working);
- broad economic change (e.g. household income, poverty levels);
- large-scale changes in attitudes, behaviour and beliefs (e.g. attitudes to work; levels and types of crime).

*METHODS:* Population censuses; regular and continuous surveys; administrative data.

### 2.4.2 Planning ahead

Social security institutions need to be able to anticipate future problems that will need a policy response so that they can plan ahead; waiting until the problem fully impacts can result in an ineffective or knee-jerk reaction. Planning ahead involves interpreting trend data to anticipate future problems and considering future policy solutions; e.g. trends may show an ageing population, but governments must consider what the policy implications might be.

*METHODS:* Economic forecasting; projections; modelling; scenario building. These methods take trend data one stage further by using informed speculation, based on stated assumptions, to build a picture of what might happen in the future.

1. This typology owes much to Nutley et al. (2002) and Weiss (1979).

### 2.4.3 Understanding a specific issue or problem

Without a clear understanding of an issue or problem, it is difficult to frame an effective policy response (or to decide that one is not needed). Good information helps to ensure that a policy addresses the problem or issue and has the desired effect and that social and economic costs are identified in advance.

*METHODS:* Surveys to inform quantification and costing; qualitative research for understanding behaviour, attitudes and understanding; case studies for understanding processes and practical issues (see Box 2.1.).

### 2.4.4 Shaping policy solutions

Once the nature of the problem is clear, a workable policy solution needs to be developed. This will depend on a good understanding of those affected, since social policy will not be effective without either coercion or public consensus. It is also important to understand attitudes and to have an understanding of those who will be affected by the policy and to "design this in". Consumers can also actively help shape policy solutions by contributing their own views and experience.

*METHODS:* Qualitative research; consultation; deliberative methods.

### 2.4.5 Understanding the principles that underlie policy

Most social policy is underpinned by assumptions about what people believe and how they behave. These are often embedded in political rhetoric, for example, "citizenship", "social responsibility" and the "contributory principle". Governments need to understand how these ideas work in practice. Policies based on misplaced assumptions risk being unsuccessful.

*METHODS:* Qualitative research.

### 2.4.6 Testing policy solutions

Rather than moving directly to full policy implementation once a policy has been designed, it may be desirable to "test" a policy in a more limited way before national roll-out. It is possible to test:

- the *policy impact* – Does it do what it was intended to do?
- the *delivery mechanism* – Does it work "on the ground" as intended?
- or both.

*METHODS:* Impact and/or process evaluation, using action and control group or random allocation design, probably drawing on both qualitative and quantitative methodologies; administrative data to monitor delivery.

### 2.4.7 Evaluating policy impact

Once a policy is implemented, it is very important to establish whether it is working as intended:

- What is the degree of take-up?
- Are there unintended consequences?
- Are there any problems with delivery?

For a policy evaluation to be effective, it needs to be planned when the policy is being designed. For the evaluation to be meaningful, the policy must have clear objectives.

*METHODS:* Evaluations are very often multi-method, drawing on both qualitative and quantitative methods. They frequently draw on administrative data.

**Box 2.1. Understanding a specific issue: Social assistance and teenage pregnancy**

- In South Africa where a Child Support Grant was introduced in 1998 to provide financial assistance towards the costs of bringing up children, there was a suggestion made in the media and in Parliament that such a grant carried with it so-called "perverse incentives":
- women would decide to have children in order to access the grant;
- teenage mothers would be particularly impacted by such "perverse incentives";
- the existence of the grant would encourage a rise in teenage pregnancy.

In two separate studies, a combination of survey and administrative data was used to investigate these claims. Neither study found evidence to support the contentions:

(1) Research undertaken by the Human Sciences Research Council, South Africa (Makiwane and Ujdo, 2006).
(2) Research undertaken by Geospace International Limited, South Africa (Kesho Consulting and Business Solutions (Pty) Ltd., 2006).

## 2.5 Making research useful

In order for research to be useful to policy, it must be credible to policy-makers, the research community and the wide group of stakeholders with an interest in social security policy. In order to be credible, research used by social security institutions must be:

### Robust

The data must be fit for the purpose. Data generated from poorly designed research – for example, badly designed questionnaires or inadequate sample designs – will be misleading. Basing decisions on poor data are risky and can lead to inappropriate or unworkable policy design. Similarly, poor-quality research which gives a misleading picture of a policy problem or a policy response is likely to attract criticism.

### Valid

Even reliable data often have limitations – for example, the research may have been confined only to a particular group, e.g. young people, a particular geographic area, or it may have addressed a limited set of questions. It is important that research is only used to inform decisions where the data are valid. In an ideal world, all the relevant data are available; in practice, this is rarely the case. Researchers need to make judgements about the limits of validity and may need to use caveats.

### As objective as possible

Objectivity is a fundamental goal of good research – never fully attainable, but always worth striving towards. Bias can come from a number of sources, for example leading questions in questionnaires or data interpretation which goes beyond what the data clearly show. If research is biased in any way, it will produce unreliable results. It will also threaten the credibility of the decisions which are based on it. It is relatively easy for a critic to identify attempts to manipulate research.

### Transparent

It is good practice to be open about research methods and research design and to publish research findings, unless there is a strong and defensible reason not to (e.g. if national security would be threatened).

Peer review is a valuable way of demonstrating transparency. Exposing research methodology, analysis and written reports to review can be a valuable way of quality-assuring research. Openness will allow others to reassure themselves that the research is robust and objective. This is important if the research is to be credible.

### Policy relevant

This is very important for research which informs social security policy. Many things are interesting, but unless research is designed, conducted and analysed in a way that reflects the concerns of policy-makers, it is unlikely to be useful. Similarly, research findings should be presented in a way that is accessible to policy-makers.

This does *not* mean that uncomfortable findings should be omitted, but it does mean that findings should be presented in a way that makes clear the relevance to policy and avoids technical jargon. Research reports often make specific policy recommendations. These must be clearly based on the findings of the research and should reflect an understanding of the policy issue. Otherwise, they will be neither useful nor credible.

## REFERENCES

Kesho Consulting and Business Solutions. 2006. *Report on incentive structures of social assistance grants in South Africa*. Pretoria, Department of Social Development. <http://www.sassa.gov.za/Portals/1/Documents/74135510-93a8-4d41-b7c6-4d438f316c9a.pdf> (accessed on 29.10.2010).

Makiwane, M.; Ujdo, E. 2006. *Is the child support grant associated with an increase in teenage fertility in South Africa?.* Pretoria, Human Sciences Research Council. <http://www.hsrc.ac.za/Document-2027.phtml> (accessed on 29.10.2010).

Nutley, S. M.; Davies, H.T.O.; Walter, I. 2002. *From knowing to doing: A framework for understanding the evidence-into-practice agenda* (Discussion paper, No. 1). St Andrews, St Andrews University – Research Unit for Research Utilisation.

Weiss, C. 1979. "The many meanings of research utilization", in *Public Administration Review*, Vol./39, No. 5.

MODULE 3

# Management of the research process

SUE DUNCAN

## 3.1 Introduction

This module discusses the various stages in the research process. For most people, the identification of an information need is the route into this process, so the discussion begins by describing the key stages in *commissioning and conducting* a research project. This is covered in section 3.2 below. Section 3.3 follows with advice on *how to manage* a research project.

Section 3.4 provides advice on the kind of *competencies* one should look for when recruiting researchers for an organization. It is likely, however, that the process of planning the project will bring up numerous other questions:

- Who can I go to for help and advice?
- What information exists already?
- Who do I need to consult?
- Where will the funding come from?

The section goes on to discuss some of these questions and to explain how an organization can adapt itself to make the whole process of procuring research and using research findings much simpler. Sections 3.4 to 3.6 cover what needs to be done to become a research-based organization. Section 3.7 discusses the different methods for conducting social security research.

## 3.2 Planning a research project: Getting started

Research managers need to be responsive to requests for research from their colleagues. It is tempting to rush straight into designing the research once the request is received; however, it is very important to spend time at the beginning of the project defining precisely what it is about. Time invested at the start of a project can help ensure that the research delivers useful outputs and that the money is well-spent.

### 3.2.1 Clarifying the research question

This is a crucial first step: What information is needed at the outset to ensure the research delivers useful outputs needs to be agreed precisely. The best way to do this is to have a discussion with stakeholders, so the researcher understands exactly what is needed and the stakeholders understand both what is possible and how their research needs are best met. It is important to ensure:

- that leading questions are avoided;
- that the research question is not set too broad or too narrow;
- that there is clarity about what question is being addressed and, in the case of policy evaluation, what the policy or intervention was intended to achieve;
- what the research output is going to be and what it will actually be used for.

This process should generate a primary research objective and probably a number of secondary objectives.

### 3.2.2 Deciding on the methodology

The first question to ask is "Is the information already available?" (i.e. "Is the research needed at all?"). If the information is not already available, the next question is "Is primary data collection needed or will secondary analysis provide the answers?"

If, after asking these two questions, it is clear that new research is needed, then it is time to ask, "What data collection method(s) is (are) appropriate?" The answer will depend on:

- the type of data being sought;
- what the data will be used for: data must be "fit for purpose". This means that it is robust enough to reliably inform the decision to be taken. Therefore, it is important to be clear how the data will be used. For example, data to be used to cost a policy option will need to have a high degree of accuracy to be useful, while data to inform an initial discussion on policy options might not need to be so robust;

- the time available: it may be necessary to make time-robustness trade-offs if data are needed urgently. If decisions have to be made at a certain time, one may have to go for the best available evidence, rather than the ideal, though caveats should always be provided if data are less than perfect; data that are so flawed that they form an unreliable basis for decision-making should never be provided;
- the budget available.

Where compromises have to be made, it is best to identify the optimum method first, so the risks of going for a suboptimal approach can be clearly calculated and spelled out to the research user.

### Box 3.1. Planning and negotiating the terms of reference (TOR)

It is important to agree a clear TOR for the project with stakeholders and to ensure that these are absolutely clear to the contractor. They should be set out as part of the research contract.

**Contracting out**

If you decide that you need to employ a contractor to conduct all or part of the work, you should:

- ensure you have identified a budget for the work;
- check the procurement rules in your organization: many organizations have in-house procurement experts and/or written procurement guidance, which you will be required to follow;
- draft a clear specification of requirements, setting out exactly what you expect the contractor to do, what information you expect to see in their tender (remember you will need to ensure they have both the resources and the skills to do the job), the required timetable for doing the work, and the outputs;
- identify suitable individuals or organizations to invite to tender. Unless there really is only one organization which can do the work, it is good practice to hold a tender competition;
- identify people to sit on the tender board: you need to include people qualified to assess the tenders from a research perspective. It is also often helpful to include a research user and someone who can advise on procurement issues.

If your organization does not have a standard research contract, you will need to produce one. It needs to reflect the special issues in research to ensure that:

- the research will be conducted both professionally and ethically;
- the confidentiality of data and data subjects is respected.

If universities or other academic organizations are commissioned to undertake the work, it is likely that they will have requirements about intellectual property rights, at the very least. For example, the university may require the right to retain the intellectual property that it brings to the project and may require the right to retain (or share) the intellectual property that arises during the course of the project.

It is good practice to promote contracts where the intellectual property is shared between the social security agency and the university. This promotes skills sharing and sustainability.

Universities will also require some rights to publish: a pragmatic solution is often for the university to be given the right to publish subject to the permission granted by the social security agency. Such permission should not be "unreasonably" withheld by the social security agency.

It is also likely that the university will require the right to use the project findings within masters' or doctoral theses.

### 3.3.3 Designing the research project

Once the objectives and methods are clear, a detailed specification for the work must be produced. This is necessary whether the project is to be conducted in-house or contracted out, so as to ensure that everyone is absolutely clear what is expected and to allow for proper monitoring of the project as it progresses. It should cover objectives, methods, research design (including sampling method), timetable, risks and how they will be addressed, analysis, outputs (data, report, etc.) and the estimated cost.

### 3.2.4 In-house or contracted out?

This is a matter of resources and skills. If the appropriate resources are available in-house, it may not be necessary to contract out the work, though sometimes contracting out can be a way of emphasizing the independence of the work.

## 3.3 The role of the research manager: Management and accountability

Once the project is scoped and plans for conducting the research are agreed, the key task is ensuring that the research delivers what is required – to time and within budget. This section describes the competencies of an effective research manager.

### 3.3.1 Research manager competencies

Every piece of research needs to have a clearly identified research manager who has day-to-day responsibility for all aspects of the project. This person must be able to act as an "intelligent customer". The research manager must have the research skills to ensure that the researcher is delivering a high-quality product and the policy skills to be able to ensure that the research is going to be policy-relevant and useful.

### 3.3.2 Communication

The key to successful research management is communication between the research manager and the researcher and between the research manager and the client (i.e. the person(s) who will be the end users of the research outputs).

The research manager must be able to liaise effectively between researcher and client and ensure each understands the role of the other. There must also be a clear understanding between all parties about what the research will deliver and when. It is the research manager's responsibility to monitor progress towards agreed targets and to take action if things are going wrong or are delayed.

### 3.3.3 Engaging stakeholders

Immediate stakeholders should be kept involved at all stages. They are the ones who are the end users of the research. Buy-in throughout the project is important.

Stakeholder management is a delicate balance between maintaining involvement and interest without making the research user feel that research is too burdensome. Some stakeholders are more hands-on, while others are content to leave many of the decisions to the research manager.

It is the research manager's job to ensure that the client is consulted on key decisions that will affect the policy relevance of the work (e.g. *topics* to be included in a questionnaire, but not the *form* of the questionnaire) and to ensure that they have mechanisms for keeping abreast of policy developments which may impact on the research.

### 3.3.4 Final reports containing no surprises for research users

It is essential that key stakeholders are kept in the picture as research findings emerge, even when they are speculative. Early findings may provide useful insights into how policies are working (or not), and early warnings of policy and delivery problems will ensure that responses to problems identified in research can be considered before it is too late.

### 3.3.5 Steering Groups and Advisory Committees

If a research project has a complex range of stakeholders with diverse interests, it is often helpful to set up a *Steering Group* or *Advisory Committee* for a project. Such a body is normally chaired by a senior manager with overall responsibility for the policy area covered by the research within the organization (often referred to as the "senior responsible owner").

It is important to make clear the power vested in the Steering Group or Advisory Committee before the project begins and before the members of the Steering Group or Advisory Committee are appointed. For example, a Steering Group may have the right to change the direction of the project as it sees fit, i.e. it "steers" the project. An Advisory Committee, on the other hand, will usually provide advice and guidance.

Steering Groups/Advisory Committees can be an efficient way of ensuring that diverse interests and perspectives are addressed in the study and of drawing on a wide range of experience. They can be a useful option to draw on a wide range of technical and scientific expertise. But it is generally advisable to limit the role of such committees to strategic issues; drafting questionnaires and research reports by committee is time consuming and inefficient.

A useful model is to have a Steering Group/Advisory Committee which advises on key issues affecting the project and a working group of technical experts who report to the Steering Group, with the chair of the working group as a member of the Steering Group/Advisory Committee.

### 3.3.6 Quality assurance and credibility

It is the research manager's job to ensure the quality of data collection, analysis and report writing and to keep the research to time and within budget. It is essential to set up processes to monitor all these elements, so there is early warning of any problems and they can be addressed before it is too late.

It can also be useful to expose research to peer review at various stages. For example, projects can usefully be peer reviewed at both tender and report stage, which can help ensure that the work is of acceptable standard but can also be an effective way of building wider credibility for the organization's research activity and exposing the organization to new ideas on data collection, analysis and related research findings.

### 3.3.7 Reporting

Reporting should *not* be something that is left until the end of the research process. It is the research manager's job to ensure that primary research users are kept fully in the picture on emerging research findings and that they are consulted during analysis and report writing to ensure that the final output is relevant and useable. This will also guard against "nasty surprises": Research will inevitably identify problems and issues that may be difficult for an organization to deal with – a policy that is not fully working as intended or a policy problem that is more severe than expected. This is a fact of life in policy research, so it is important that senior managers and politicians have an early warning, so that solutions can be considered in good time.

The research manager also has a very important job in protecting the integrity of the research report. This is really an extension of the quality assurance role and involves ensuring that the reported research findings and recommendations are a true reflection of data and analysis. Peer review can also be an important way of strengthening this process.

### 3.3.8 Publication

It is good practice to publish research findings, unless they jeopardize the right of individuals to confidentiality (e.g. the results cannot be made anonymous or threaten national security), although such cases are likely to be exceptional. Wide dissemination encourages policy debate and demonstrates that organizations are willing to expose their thinking to wider scrutiny.

It is advisable to develop a research publication policy within the organization: this helps to minimize last-minute problems with publication of sensitive research findings. It is also advisable to consider dissemination at an early stage. A provisional dissemination plan should be produced in the early stages of the project and regularly reviewed throughout the life of the project.

### 3.3.9 Research utilization

Publication will allow research findings to reach a larger group of stakeholders, but it does not alone ensure that the research outcomes are actually *used* within the organization. Planning and follow-up on this are yet another very important part of the research manager's role.

Research outputs must be produced with the user in mind. The research manager must actively assist research users to apply research to their area of work. This will involve explaining the relevance of findings and assuring the robustness of the research.

Research utilization is an ongoing process. Close user–research manager relationships, forged at the beginning of the project and maintained throughout, are the best way to ensure that research findings are given serious consideration, even when the findings are uncomfortable.

Trust is essential in the relationship between the research user and the research manager. Without trust, there is a risk that research findings will be ignored. Inevitably, the research manager will have a closer knowledge of the research findings than the user does, and it is part of the manager's job to play a continuing role in working with the user to ensure relevant research findings are fully considered and understood in organizational decision-making.

## 3.4 Setting up a research-based organization: Identifying existing resources

Social security research and researchers are part of a wider research community. To use research effectively and to make optimum use of resources, it is necessary to link up with this community. This section describes the research resources already available and explains how to access them, while section 3.5 explains how organizational systems and processes to support and encourage a research-based approach can be used.

### 3.4.1 Mapping the research community

#### Subject experts

Identify subject experts in universities, research institutes, pressure and interest groups, trade and professional organizations. They are people who conduct research and study and analyse poverty, social assistance and related subjects.

### National and international partners

Persons in one's own country as well as those in other countries and international organizations with relevant expertise should be included. By collaborating with them, one can maximize the benefit of advice and expertise received from external partners.

### Communication networks

Map out and set up new communication networks. Check the scope for cooperation, knowledge sharing and expertise pooling.

## 3.4.2 Mapping research technicians

Identify those who can *conduct research*. The list will probably include some of the experts identified above. These will be commercial survey organizations, qualitative research organizations, independent researchers, academics and research institutes.

## 3.4.3 Sources of advice

The challenges faced in setting up a research capability will have already been faced by others – it is important to make the most of the advice available, in order to avoid costly mistakes or reinventing the wheel. Other ISSA members or development-oriented research organizations such as the Overseas Development Institute (ODI), for example, may be willing to make their expertise, experience and partner networks available below "market rates".

## 3.4.4 Mapping internal expertise

Internal analytical expertise is very important. People with a good knowledge of the organization and its policy priorities, as well as analytical skills, are needed to manage the policy-research interface.

## 3.4.5 Intelligent customers

Research must have an "intelligent customer" to act as a bridge; if not, it will be difficult to make full use of the research. This person will need to identify and refine research needs, to manage research projects, and to help the organization interpret and make use of research findings.

## 3.4.6 Research partners

It is also important to know who else within the organization has research or analytical skills so that one can make the most of all available services. These might include actuaries, demographers, statisticians and economists, as well as people with specific social, political, historical or anthropological research expertise. Try to identify not only partners with technical skills but also partners with knowledge, views and good judgement.

## 3.4.7 Mapping existing knowledge and data

Before the organization's research requirements can be identified, one needs to know what information is already available. This will avoid wasting money on researching issues that have already been researched.

One first needs to identify what is available in the various registers and databases of the organization – statistics, demographic and administrative data – and in other monitoring documents (such as routine returns) as well as in other research projects in progress or completed.

Next, external sources must be identified. Again, these might be statistics or other monitoring returns or knowledge from completed research.

### 3.4.8 Commission an expert

The most efficient way of mapping existing knowledge may be to commission an expert or experts to conduct systematic reviews in key policy areas.

### 3.4.9 Setting up knowledge management systems

As well as mapping what is known already, systems must be set up to keep the organization's knowledge up to date (see section 3.4.1 "Mapping the research community").

The organization's internal research function should include a documentation and research intelligence service so that the social security institution can draw on relevant research insights, wherever they are conducted.

**Box 3.3. What does a good social researcher do?** [1]

**Policy and delivery focus**

- Understands and directs effort to meet customer needs.
- Works in partnership with other analysts, policy colleagues and broader customers to provide relevant and high-quality contribution that adds value to government policy decision-making.

**Delivering results**

- Plans work activities, reviewing and prioritizing as necessary, to achieve high standards and meet deadlines.
- Is proactive and uses initiative when problems arise or progress is slow.
- Shows resilience under pressure and does not let setbacks affect performance.

**Learning and improving**

- Acknowledges own development needs and seeks new skills, knowledge and opportunities for learning.
- Learns from others.
- Adapts quickly and effectively to new people, situations and task demands.
- Operates effectively in a range of roles and contexts, including times and situations of uncertainty.

**Critical analysis and decision-making**

- Critically evaluates data and information with accuracy and perception and is able to synthesize and use data drawn from a variety of different methods appropriately.
- Makes sound, evidence-based decisions (and/or helps others do so).
- Assesses risks and defends decisions and actions.
- Responds effectively to unforeseen situations.

**Constructive thinking**

- Thinks imaginatively while keeping the goal in mind.
- Understands the bigger picture and can make the link between issues.
- Shows an open mind with the intellectual rigour to generate original ideas and develop practical solutions from them.
- Is able to facilitate, encourage and build upon the ideas of others.

1. *The Social Researchers' Competency Framework* of the UK Government Social Research (GSR) Unit is a good example of a comprehensive set of competencies relevant to research work and grades in public service. It is designed for use in recruiting, promoting, appraising and developing social policy researchers. See: http://www.gsr.gov.uk

**Box 3.3. What does a good social researcher do?**

**Professional expertise**
- Demonstrates the detailed knowledge and experience necessary for the job of a government social researcher, expressing the core technical capability, knowledge and awareness in terms of behaviours.

**Developing constructive relationships**
- Uses interpersonal and other communication skills to build rapport with others.
- Shows awareness of the effects of own behaviour on others and understands their situations and concerns.
- Values diversity and shows flexibility of style.

**Communicating with impact**
- Communicates written and oral information clearly, concisely and persuasively.
- Communicates own viewpoint succinctly and defends it appropriately.
- Facilitates discussions effectively to achieve clear outcomes.

**Leading and directing**
- Takes an active and prominent role in providing direction to staff and contractors.
- Champions high standards.
- Gains the trust, commitment and cooperation of others.

## 3.5 Setting up a research-based organization: Systems and processes

Once knowledge and expertise have been mapped, it is important to establish ways of ensuring that the organization identifies the information it needs, obtains that information and feeds it back into organizational decision-making.

### 3.5.1 Research planning: Identifying strategic research needs

A system for identifying and prioritizing the organization's strategic research needs on a regular basis should be set up. This should ideally dovetail with the organization's planning cycle. Research priorities should reflect organizational priorities, so that when the organization's management board (or equivalent) agrees business priorities it should be possible to identify at the same time the organization's high-level information needs. These might be as broad as *understanding the extent and nature of poverty and social insecurity* or *reaching those engaged in the rural and informal sectors*. The detailed research questions will be defined later.

### 3.5.2 Research planning: Trends and future challenges

It is important to look beyond immediate needs. The research planning process should consider the need for regular data to enable monitoring of trends and data needed to inform decision-making in the shorter and medium term. There should be a process for anticipating future information requirements. This would involve scanning new data and reviewing trend data to identify emerging problems and issues which may require a policy response in future.

Collecting reliable data can be a time-consuming business, so it is important to plan ahead and to make maximum use of the regularly updated data available from the administrative registers and databases of the social security administrations.

Once a problem hits the policy arena, it may be difficult to collect data in time to make a useful contribution to policy formulation. Policy solutions cannot always wait for research data to be available.

### 3.5.3 Consulting stakeholders

Consulting stakeholders is an important part of the research planning process. A regular means of consulting external stakeholders needs to be set up. This may be through conferences, workshops, newsletters or a combination of these.

The *primary stakeholders* will normally be people within the organization, since they will be responsible for delivering the organization's work programme.

Policy stakeholders will have views on their information needs, but the needs of practitioners and others in the delivery chain may be different, and it often makes sense to address them at the same time, rather than in separate pieces of work. Similarly, consumer groups will have their own priorities.

Academics and others in the wider research community will also have useful views; organizations tend to focus predominantly on their more immediate priorities, while researchers and academics may be the first to identify emerging and longer-term issues in order to make connections across organizational boundaries and to highlight unanswered questions identified in existing research.

Other research funders can also make an important contribution to research agenda setting. Organizations with similar or related priorities may have different but useful perspectives on the same issue. They may have a joint interest in the organization's research priorities and may be willing to work and/or fund research jointly.

Ultimately, it will be for the organization itself to decide on its priorities, but a culture of openness helps to give credibility to an organization's activities, encourages dialogue between wider research and policy and maximizes the opportunity to look beyond organizational boundaries and time horizons.

### 3.5.4 Other ways of encouraging a research-based organization

The research planning process is one way of encouraging an organization to review its research needs. The more closely the process is tied to organizational planning processes, the less likely it is that research needs will be overlooked. But there is still a risk that research needs will not be addressed, because policy-makers are "too busy" or think they have all the answers or believe that politicians have made their wishes clear and research cannot help.

Processes which require policy-makers to set out what evidence they have consulted in developing policy can be an additional driver. Of course, the decision may ultimately be influenced by political, financial or organizational considerations, but politicians and other decision-makers need to have the opportunity to review available research evidence in coming to a decision. This will help ensure that policy-makers focus on information needs at the data collection stage and in feeding research back into the policy cycle when results are available. There also needs to be a process for regularly reviewing the research programme, so that it can adapt to changing organizational priorities and needs.

### 3.5.5 The management of knowledge

The management of knowledge which data unlock is as important as the collection of data. Good data and analysis are not useful unless they are accessible.

Reviewing the evidence available to inform policy decisions is an ongoing process. New research evidence needs to be weighed against existing research knowledge, so that the insights available to policy are constantly updated.

It should be accumulated bodies of knowledge which are brought to bear on the decision-making process. The findings of individual projects are only useful if it is clear how they challenge or illuminate existing knowledge. It is the job of the researcher to make sure research is available in this form. It is the policy-maker's job to take ownership of the knowledge base which informs his or her policy area.

## 3.6 Setting up a research-based organization: Funding and collaboration

Reliable data collection and analysis are expensive. So, if value for money is to be secured, it is crucial that research focuses on organizational priorities and that it is professionally managed. Otherwise, money may be wasted by doing the wrong research or doing research that is of poor quality or irrelevant.

### 3.6.1 Joint funding

Thorough stakeholder consultation may well identify areas of common interest where there is scope for joint funding. This can be a very cost-effective way of conducting research, but the process must be carefully managed.

### 3.6.2 Working together

It is essential to have project management arrangements and lines of responsibility to set out clear terms of reference, define how decisions will be taken and agree practical issues such as data access and ownership arrangements. Without such agreements in writing, there is a risk that funding issues will hamper the effective conduct of the research.

### 3.6.3 Non-cash funding

Contributions may be "in kind" – for example, staff resources (expert advice or training) – or cash. Whatever the method, clarity about responsibilities and arrangements is essential.

### 3.6.4 Liaising with other sources of funding

Strong and effective liaison with other sources of funding of research may enable the organization to influence wider research agendas (e.g. national research councils, universities and so on), either by suggesting fruitful research topics or additional areas to be covered in research projects.

Good liaison across funding organizations can reap clear benefits for all sources of funding and can help to ensure that scarce funding resources are used optimally.

### 3.6.5 Data sharing

Similarly, good stakeholder mapping and/or identification of experts and data sources may provide opportunities for data sharing,[2] thus saving duplication of effort. The terms of the arrangement and the rights of the parties involved must be clearly set out.

### 3.6.6 Earmarked research funds

Within the organization, there needs to be an earmarked budget for externally commissioned research. Ideally, this should be held and managed at a strategic level in the organization. Responsibility for oversight and allocation of funds should be clearly identified. Criteria for budget allocation should be properly defined, reflecting organizational priorities. Localized budgets tend to work against a strategic approach to research priorities. There is a risk that overall organizational priorities could be distorted by parts of the organization with the most influence and/or resources. Localized research budgets also make it more difficult to reallocate resources as organizational priorities change.

2. Of course, protecting the confidentiality of the individual is important.

### 3.6.7 Budget flexibility

There also needs to be a process for a regular review of the organization's research needs as priorities change or new priorities emerge. The management of the research budget should be sufficiently flexible to deal with such changes which may well emerge later. For this reason, it is prudent to release funding on a rolling programme throughout the year and to have a contingency element in the budget.

## 3.7 An introduction to different methods for undertaking social security research

Research methods are usually broadly divided into quantitative (sometimes called statistical) methods and qualitative methods (some qualitative approaches, particularly observational techniques, are referred to as "ethnographic" methods). Neither approach is "better" than the other, and each has important contributions to make to evidence-based social security policy-making. They serve different purposes, and there is a growing appreciation of the need to incorporate both in research design.[3]

**Quantitative** approaches usually involve:

- data collection through cross-sectional or longitudinal surveys or from administrative data;
- analysis using statistical analysis software such as STATA, SPSS, EXCEL and SAS;
- simple bivariate (e.g. cross-tabulation) analysis or more complex statistical or econometric modelling.

Quantitative research methods may be used to:

- assess needs and forecasting and actuarial modelling;
- monitor delivery of policies;
- evaluate policies;
- assess opinions or attitudes through social attitude surveys in order to inform policy design or improve service delivery.

**Qualitative** approaches usually involve:

- data collection through in-depth interviews, focus groups or through participant or non-participant observation;
- analysis using, for example, computerized textural analysis software packages such as NVivo.

Qualitative approaches may be used to:

- assess needs;
- evaluate policies; and
- assess opinions or attitudes in order to inform policy design or improve service delivery.

3. Indeed, there is an international movement to promote this combined approach known as "Q-squared" or "Q2", based at the Centre for International Studies at the University of Toronto in Canada – see http://www.q-squared.ca. This Centre "aims to promote a better integration of 'qual' and 'quant' approaches to poverty assessment in the developing world through information sharing and networking, national capacity building and the piloting of Q2 methodologies".

MODULE 4

# Statistical research on social security

FLORENCE BONNET – THEO BUTARE – AIDI HU

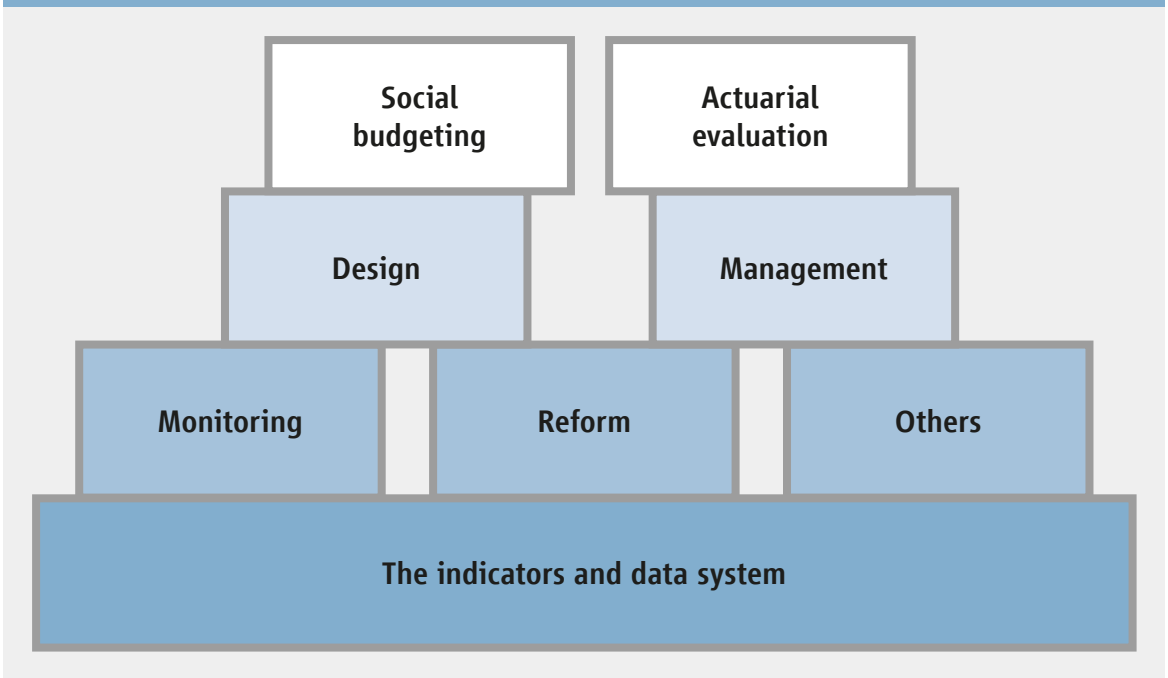
## 4.1 Statistical research on the current performance and impact of social security programmes

Social security-related data and indicators are permanently sought for various purposes: the insured want to know how much they have to pay in terms of contributions and how much they will receive in terms of benefits. Employers need to know the overall amount of contributions due.

Social partners, interest groups, academic researchers, policy-makers, managers and monitors of social security schemes require a more comprehensive database to be able to examine and improve the governance of social security as a whole.

For achieving good governance, appropriate indicators and accurate, up-to-date and complete data are indispensable. Without them, the design, management, monitoring and reform of social security systems will not work properly. Therefore, the indicators and data system can be considered the primary quantitative foundation of good governance in social security (see Figure 4.1.).

Figure 4.1. The indicators and data system are the quantitative foundation of the governance of social security

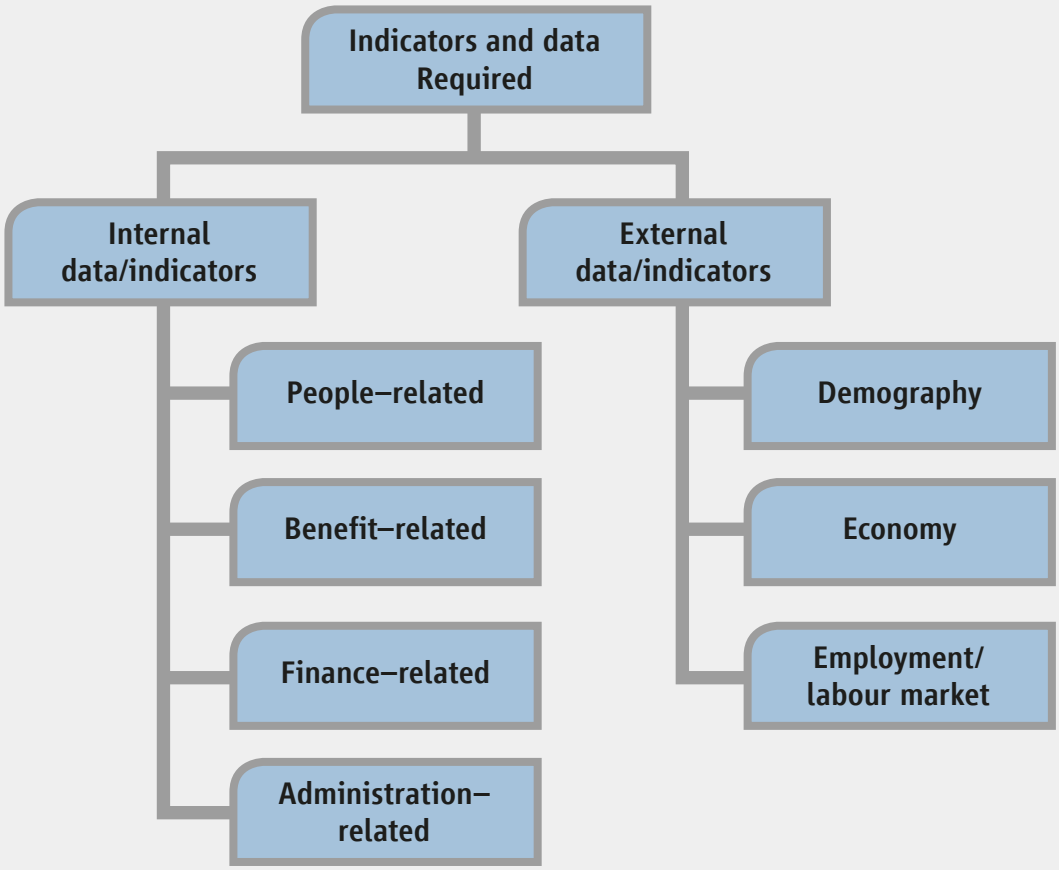



It is highly advisable for social security administrations to gather statistical data on their schemes as part of their regular operations. This is often a great lacuna. In many countries, data collection for actuarial valuations starts "from scratch" each time there is a valuation, which is unnecessary and a waste of resources.

A national social security scheme has (or should have) the largest database in the country. The national statistical office can also use the social security scheme statistics. For example, good wage statistics from social security scheme data can often be better than statistics based on labour force surveys.

The choice of correct indicators and the quality of the database are vital issues for social security administrations. For the purpose of governance, both internal (scheme-specific), and external indicators and data are needed (see Figure 4.2.). Only the internal indicators and data can be directly influenced by the managers of social security schemes.

Figure 4.2. Categories of data/indicators




Detailed quantitative information to be contained in each of these four sub-categories of the internal data/indicators includes:

- People-related:
  - beneficiary by sex, age, income, social group, occupation/sector, work contract, region, etc.;
  - coverage rates: actual coverage, legal coverage and/or potential coverage;
  - insured persons as or not as contributors by sex, age, earnings, social group, size of establishment, occupation/sector, region, etc. and coverage rates (actual coverage against legal and/or potential).
- Benefit-related:
  - average amount of benefit granted by sex, age, social group, occupation/sector, region, etc.;
  - level of benefits in terms of reference wages;
  - adequacy of protection in terms of poverty prevention;
  - distribution in terms of benefit disparity;
  - impacts on employment/labour market.
- Finance-related:
  - current total revenue and revenue by sources (contributions, subsidies, investment income, etc.);
  - investment and rate of return by market instruments, sector, domestic/international investment;
  - current total expenditure and expenditure by benefit, administration, etc.;
  - administration costs as a percentage of revenue, expenditure or their combined total;
  - current and accumulated balances and balance rates (balances as a percentage of expenditure or revenue);
  - investment structure and annual return rates by instruments and by domestic/international markets.

- Administration-related:
  - quality of benefit delivery, such as the length of claim processing, timeliness and accuracy of benefit payment;
  - general services, such as client-oriented information/communication system, procedure of registration, contribution collection and benefit claim, etc.;
  - fraud inspection and compliance enforcement;
  - complaints and appeals raised and settled;
  - human resource development in terms of staff capacity by sex, age, education, training, ratio of staff to client, ratio of staff salary to reference remuneration, etc.

Section 4.2 will cover various issues on how to build up and manage a reliable and well functioning database, such as:

- data sources and types;
- availability and quality of data;
- techniques for designing, collecting, storing and processing data; and
- analysis methods.

## 4.2 Broad categories of data needed for statistical social security research

The objective of this section is to present several of the broad categories of data required for research in the field of social protection and discuss the relevance and limitations of the principal sources of quantitative data on social security in relation to the information sought. This discussion is complementary to the following section (4.3) on "Possible data sources".

The generic categories of data, as well as the more specific sources of quantitative data on social protection, apply to data required for the management of social security systems, for actuarial reviews of social security systems or for determining the performance indicators for social security systems. They also apply to data used within the evaluation framework of social security policies, some of which are obtained through aggregating or consolidating data collected at an individual scheme level. In efforts to extend social security coverage, the data enable an analysis of the population currently not yet covered by social security.

### 4.2.1 Data specific to the scheme and data produced by external sources

*Data specific to the scheme* are collected mainly as part of the ongoing activity of the scheme. Data in this category relate specifically to the population covered (contributor, member or beneficiary). In the case of contributory systems, data can relate to the contributors' income and the contributions paid. The data can also cover the benefit expenditures and administrative costs, investments and assets. These data are known as "administrative" data. Legislative data concerning the scheme are also included in this category.

Data produced *by sources external to the scheme* include at least three categories of data which are useful for analysing social security systems and which help in analysing social policies:

- *social security data on a national or international level.* Data in this category can be statistical data, **quantitative** data (social security expenditures or revenue) or **qualitative** data on legislation, such as national legislative provisions and international Conventions;
- *contextual data*, which do not include data on social security but which are used for analysing social security systems and/or social policies: denominators of the indicators, evolution assumptions in projection models and, more broadly, for contextual analysis and its impact on the social security system as a whole or on a particular scheme. Data in this category comprise demographic, economic and social data;

- *information on the living and working conditions of the population currently not covered by social security.* The specific objective of this information is to identify the different groups not covered by social security and to evaluate the risks which these groups face and the requirements and strategies to overcome these risks.

### 4.2.2 Qualitative data and quantitative data

In undertaking quantitative analysis and research on social security, certain **qualitative** information and variables are required:

- directly, as coded qualitative variables which are included in the analysis;
- indirectly, as complementary information to guide the analysis, to help in identifying the modelling parameters for projections or to interpret the results.

There are at least two definitions of the term "qualitative data":

- One definition is a body of information in the form of a *narrative* such as information collected through interviews or focus groups. Information in this form cannot be used specifically for statistical analysis but can be used as supplementary information to quantitative data.
- Another definition is *categorical* data. This definition is used to describe qualitative variables. As the name suggests, qualitative variables contain values which express a "quality" or "attribute". This is particularly true of coded variables which describe, for example:
  - the type of social security programme (social assistance, social insurance, employer responsibility, universal programme, etc.);
  - the function (old-age, invalidity, survivor, unemployment, occupational accidents and sickness, etc.) of the scheme; and
  - the categories of the population covered by a specific programme (private sector employees, independent workers, public sector employees, agricultural workers, etc.).

Finally, qualitative variables can be *nominal* (as in the above examples) or *ordinal*, where the methodologies describe a preference order or classification such as "1. Decreasing; 2. No change; 3. Increasing".

Unlike qualitative variables, **quantitative** variables include *measurable* values. Many variables in this category are used to analyse the performance of social security systems and, in particular, to construct monitoring and performance indicators (made up of one or more variables) or within the framework of a modelling exercise on projections. Quantitative variables can be continuous or discrete.

A quantitative variable is said to be *continuous* when the methodology is expressed on a continuous scale of values as, for example, in social security expenditure or expenditure as a percentage of GDP. *Discrete* variables can only have discrete values – in other words, values which are represented by whole numbers as, for example, in the number of beneficiaries in a particular scheme.

Quantitative data can be useful, for example, in certain studies to record the variables in new categories for the purposes of statistical analyses.

### 4.2.3 Primary data and secondary data

*Primary data* are data observed or collected first hand. They include data collected through household surveys, censuses, interviews or group discussions with those directly concerned. Household surveys which are regularly undertaken (labour force surveys, surveys on household living conditions) include questions on coverage – particularly, health and pensions coverage – as well as the amount of social security benefits received. *Secondary data* are published data (such as annual reports published by institutions or reports on surveys) or data previously collected or collected by other institutions. In other words, secondary data are existing primary data previously collected by a third party or for other purposes.

The collection of primary data is more costly and requires significantly more resources than the collection of secondary data, but it is, nevertheless, essential from time to time. Section 4.4, "Main techniques of collecting, storing and processing data", describes several data collection techniques.

### 4.2.4 Administrative data versus household survey data

Administrative data are collected as part of the current activities undertaken by social protection institutions. They are fundamental data which enable the continuous monitoring of a scheme's activities with the help, in particular, of monitoring and performance indicators. Administrative data are "raw data" used to analyse a scheme's previous or current performance and are the reference point (reference year and previous trends) for building *assumption models* to project future short-, medium- and long-term demographic and economic trends.

Household surveys also represent a source of statistical data on social protection. However, they are largely underutilized in developing countries. They represent a valuable source of information for estimating the coverage of different benefits (particularly health and pensions) or different social policy programmes. The comparative advantages and respective limitations of these sources are addressed in the next section.

Census data provide a third source of information, mainly on the demographic structure of the population and potential beneficiaries.

## 4.3 Possible data sources – which data for what kind of indicators and objectives: Pros and cons

This section considers the use of administrative data and household surveys as data sources for social security policy research. Both types of data can be very useful. Yet, in both instances it is important to take into account the quality of the data and the extent to which the data adequately address the policy questions that are being considered as part of the research exercise.

### 4.3.1 Administrative data

In this context, administrative data are data collected by the social security administration about its beneficiaries. They are routinely collected as part of the assessment and payment administrative procedures. Such data are therefore not specifically collected for research purposes.

Information concerning beneficiaries, benefits and the covered population is recorded in order to set up detailed information on the scheme's revenue and expenditure, to manage the system of services, to monitor given services and to ensure quality.

In Europe and in particular Northern Europe, administrative data have become an important source of information for research. Population registers are routinely constructed from identity document information and linked to social security and other administrative data, such as from education and health contexts, to produce comprehensive data at an individual level which is in effect a census of the whole relevant population. This can be undertaken cross-sectionally (for a single time point) or longitudinally (by linking individuals at several time points).

These techniques of using administrative data have become increasingly prevalent in developing countries (see Box 4.1. about the International Data Forum's (IDF) findings about administrative data in China, Brazil, India and South Africa). A number of initiatives are in place to enhance the quality of administrative data and to explore ways in which such data can be used.

Administrative data may be collected at the central, regional and local levels, although in the latter case unified data standards are necessary to ensure regional data comparability.

**Box 4.1. The International Data Forum**

The International Data Forum (IDF) was established to:

- provide mechanisms through which data needs for future cross-national collaborative research on social scientific issues can be identified and prioritized;
- coordinate efforts by national research funding agencies and statistical authorities to make data more widely available for research purposes.

For the first international IDF Conference in Beijing in 2007, social science microdata-scoping studies were undertaken for Brazil, China, India and South Africa. These four documents and additional details about the IDF can be downloaded from http://www.internationaldataforum.org.

*Advantages of administrative data*

- Such data are unintrusive to collect (though measures need to be taken to ensure that the researcher has the right to use the data);
- There are few additional costs in making the data available, as the data set is already in place (unlike a household survey, which is very costly to implement);
- They are comprehensive in that all recipients are contained within the data set.

*Drawbacks of administrative data*

- Administrative data usually contain ample information on those groups of the population that are covered by social security, but not on those who are *not* covered;
- While administrative data can be used to estimate the extent of coverage, they usually do not provide any insights into the causes and effects of non-coverage;
- Eligible non-recipients are not captured;
- It is necessary to guard against double counting if it is not possible to identify individuals using a unique reference number or ID when more than one benefit is involved.

### 4.3.2 Sample household surveys

Sample household surveys are designed to understand the situation and behaviour of individuals and households and are particularly useful for welfare analysis and for impact analysis of a given social security programme.

*Understanding individual and household decision-making.* Such surveys are typically based on complex and detailed questionnaires, covering the whole population but with a relatively small sample size. They offer a host of information on the socio-demographic and socio-economic characteristics of individuals and the households in which they live, including:

- household size and composition;
- demographic characteristics (age, sex, ethnic group, nationality, migration etc.);
- employment status;
- wages and other income;
- social security benefits received; and
- household expenditure and sources of expenditure.

*Monitoring welfare indicators.* In some countries – especially where complex household budget surveys are not available – large-scale rapid monitoring surveys are conducted. They are designed for the specific purpose of social exclusion and social protection evaluation and cover a more limited set of data. The major drawback of such surveys is that, due to information restrictions, causal analysis at the household level is rarely authorized.

*Impact evaluations* aim to reveal which impacts were caused by the social transfer and which might have been caused by other factors. Survey data allow for an efficiency and effectiveness analysis of different benefits, especially for assistance-type programmes (means-tested benefits, conditional cash transfers).

Types of surveys include:
- population and agricultural censuses;
- labour force surveys;
- household budget surveys;
- income and expenditure surveys;
- living standards measurement surveys;
- priority surveys;
- core welfare indicators surveys;
- demographic and health surveys; and
- multiple indicators cluster surveys.

These surveys differ in terms of the goals they serve, population coverage and the scope of questions. For the purpose of evaluating social security, the most significant are surveys other than censuses.

### 4.3.3 What can survey data be used for?

Household survey data are particularly useful for estimating the following measures that can aid social security monitoring:
- gathering information on the distribution of specific risks among the population, correlates of specific risks, and participation in social security;
- identifying existing gaps in coverage and analysing factors related to these gaps (e.g. where social security is not reaching specific social and economic groups or regions). This is often best undertaken in conjunction with administrative data;
- calculating aggregates of individual or household income and expenditure that allow for the estimation of insurable earnings, average wages and average expenditure. In some cases, where information on taxes and contributions incurred is available and reliable, the aggregate level of taxes and contributions can be assessed. Listed aggregates are often used as denominators for the calculation of specified performance indicators (e.g. average replacement ratios of benefits in payment, administrative costs in relation to total insurable earnings);
- calculating aggregates of the total employed population, total insured population and total beneficiaries that are used to estimate scheme indicators (e.g. scheme demographic ratio);
- evaluating the effectiveness of schemes in terms of outcomes achieved (e.g. increasing levels of income in relation to the poverty line, the impact of social protection on education and health outcomes);
- modelling the performance of social security schemes in the long term; and
- particularly with respect to coverage and benefit levels as well as to the impact of external factors (e.g. increasing/decreasing unemployment, increasing consumption and levels of income).

### 4.3.4 How is survey-based research undertaken?

Once the objectives of the research have been clarified, the main steps are to define the appropriate methodology and data collection tools (questionnaires, sampling methods, reference population, etc.) to collect data, also often referred to as "fieldwork"; data entry; and verification before analysis.

### 4.3.5 Drawbacks of sample household surveys

The scope of information on social security varies between different countries. In many cases, even though information on benefits received may be available, information on contributions is not sufficient. Thus it is often not possible to properly evaluate the number of contributors to different schemes.

Another drawback in using survey data to assess social security coverage is the impossibility of distinguishing between new beneficiaries and the stock of beneficiaries.

Other problems include: cost (household surveys can be very expensive); response bias (low-income people are less likely to participate in surveys); the difficulty of designing sample frames from which to draw representative samples (particularly in developing countries which may not have electoral registers or geo-coded address files, etc.); respondent fatigue (some deprived areas feel "over researched"); intrusiveness (very detailed questions are often asked of the interviewee); and difficulties of recall, especially for detailed income/expenditure questions. There is also the risk of under-reporting of social assistance receipt.

Generally, surveys conducted in developing countries in recent years address a much broader range of information regarding social security than before. Thus, in time, if surveys are conducted on a regular basis, the information base on social security will be broader and the potential for using surveys to evaluate social security will increase. Therefore, sample household survey data are a rich source of information for undertaking social security policy research which has not yet been tapped to the full.

Similarly, administrative data have been underutilized for social security research in developing countries, but this is undoubtedly an area where rapid progress is being made.

**Box 4.2. Using administrative data: Research on take-up of social assistance benefits**

For the social security system to function properly, it is essential that those eligible for assistance are not impeded from receiving it, that reliable data are available on take-up rates and that research is carried out to identify the factors that influence take-up. A key aim of this project was to provide a valuable information source which allowed the effectiveness of the social grants programme to be assessed.

The South African Taylor Committee Report into the Social Security System published in 2002 reported estimated take-up rates of 90 per cent for the Disability Grant, 85 per cent for the State Old-Age Pension and only 20 per cent for the Child Support Grant. Through effective targeting, the Department of Social Development wanted to maximize take-up among those eligible for social security assistance.

Using a combination of geo-coded administrative data on grant recipients (for the numerator) and census data (for the denominator), eligibility and take-up rates of the Old-Age Grant and Child Support Grant were calculated and mapped at municipality level to identify areas where take-up is low. The work was used by the Department of Social Development to focus its efforts on promoting take-up.

Research was undertaken by the Centre for the Analysis of South African Social Policy at the University of Oxford (see Noble et al. 2005, 2006).

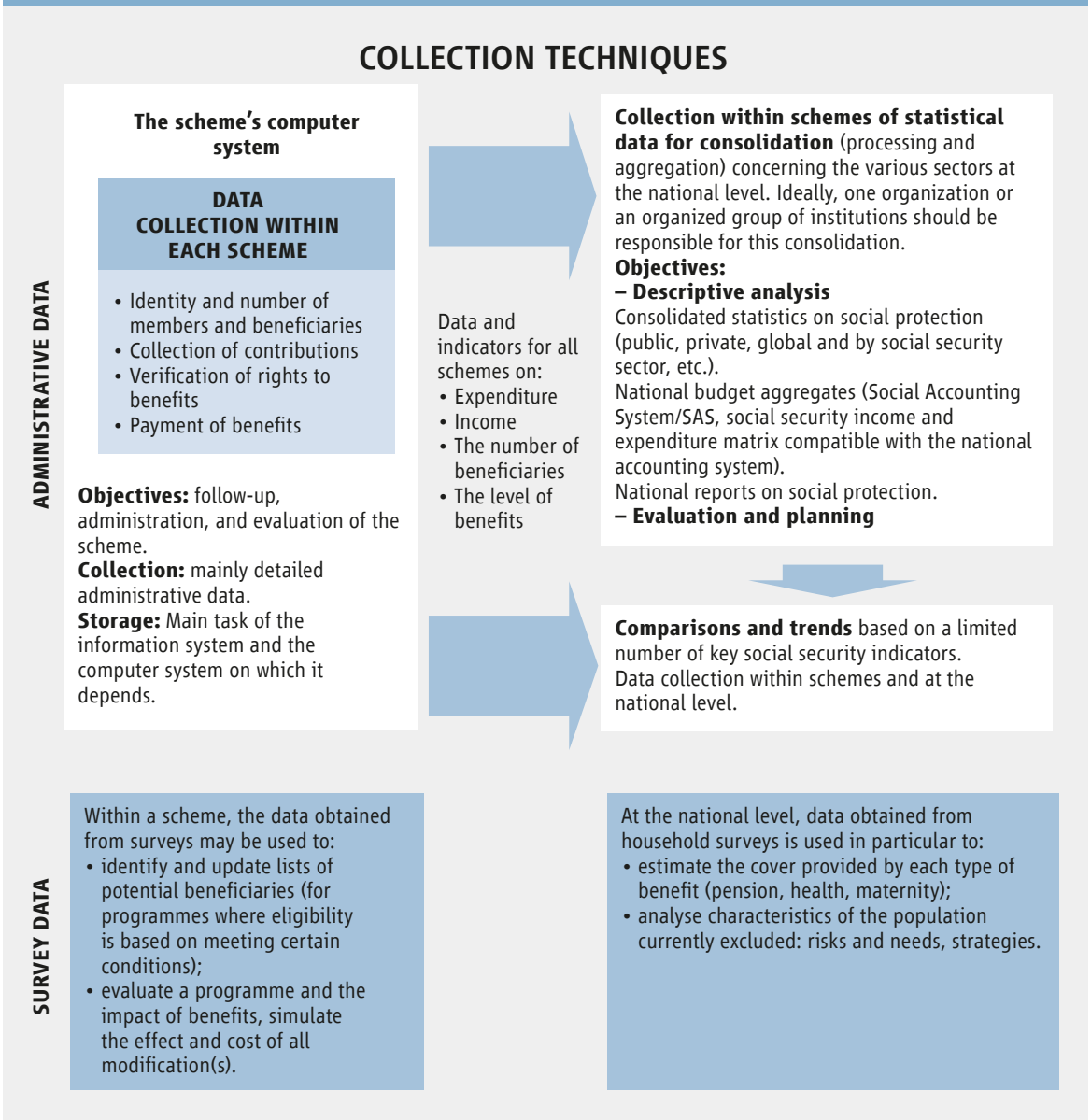
## 4.4 Main techniques of collecting, storing and processing data

The techniques for collecting, storing and processing data vary, depending on the users and their objectives. This section distinguishes between *scheme administrators*, whose main aim is to ensure the efficient management of the activity, and *outside users*, in particular analysts and political decision-makers, researchers and institutions at both national and international levels, who are more interested in descriptive analysis or analysis for policy-making.

The first section provides information on the collecting and storage of data within schemes. The second deals with data at national and international levels.

Figure 4.3 shows the main levels in the collection of data on social protection. The schemes and the objectives of their administrators are shown on the left; outside users are shown on the right.

Figure 4.3. Levels of data collection, objectives and main techniques



### 4.4.1 Information concerning the collection, storage and processing of data in order to follow up the activities of the scheme

The data collected and processed at the level of each scheme are raw data, providing a "foundation" for follow-up of the activities of the scheme as well as reference data for actuarial studies. This section deals mainly with the data required for follow-up.

Follow-up can be defined as a continuous activity which consists of: keeping track of the progress of planned activities and providing timely information for good management and effective decision-making.

The follow-up system includes all the procedures, tools, information networks and tasks which enable the collection and processing of the data obtained as part of the follow-up process.

### 4.4.2 Principal data required for ongoing administrative follow-up of a social security scheme

Certain data and collection techniques are common to all systems; others are confined to only some of them, depending on the nature of the scheme in particular, such as:

- the distinction between contributory and non-contributory systems;
- the method of selection of beneficiaries, including the distinction between universal coverage (of the entire population or a given population group) and defined target groups on the basis of means-testing or the behaviour of the beneficiary (mainly in terms of education and health).

Other factors which may also be involved include the identification of beneficiaries, their selection and verification of eligibility. The latter may depend on national legislation, on the type of membership (compulsory, voluntary or automatic) or on the public or private nature of the scheme.

### 4.4.3 The basic data required by all institutions responsible for the administration and follow-up of the activities of social security schemes on a regular basis

The administration must have access to up-to-date information concerning the identity and number of beneficiaries (sex, age and, if appropriate, family status and employment situation) at all times.

In terms of policy evaluation and design (particularly as regards extension of coverage), it is also important to have information on the identity, characteristics and needs of potential beneficiaries who should or could be covered but are excluded.

Beneficiaries' rights should be monitored, whenever necessary, as well as which benefits should be paid, how much and when, and how much has already been paid.

The expenditure involved in the payment of benefits including the type and amount of other administrative costs should be calculated.

Benefits should be paid to claimants and/or, in the case of health insurance, to service providers.

In the case of contributory schemes (including micro insurance), the data must also enable the administration to access information concerning the number of members and active contributors at all times, as well as the amount of outstanding contributions, and to collect outstanding contributions.

### 4.4.4 Main collection techniques

There are two main types of social security data and techniques for collecting them:

- *Administrative data* collected in the course of the day-to-day operations of a scheme, either on paper or computerized (particularly for the registration of beneficiaries) or through the transfer of data between several sectors of a computer-based system or from external databases (e.g. between the Benefits and Payment of Benefits sectors). These transfers may or may not be automatic;
- *Survey (or census) data* collected via a questionnaire or a guided interview. These primary data are collected directly from households/individuals. The completion of a survey or census involves a certain number of steps and activities, including:
  1. *The statement of objectives:*
     - clearly defined objectives;
     - identification of the users of the results and their needs;
     - inventory of existing information;
     - inventory of the human and financial resources required and available;
     - available budgetary resources and costs;
     - time limits.
  2. *Technical characteristics:*
     - sampling method: type of survey, target and observed population, census base, size of the sample, etc.;
     - questionnaire: type of questionnaire (open-ended and/or closed questions, subjects covered, etc.);
  3. *Implementation, fieldwork and analysis:*
     - collection: testing the questionnaire, survey timetable, staff and its responsibilities, training of researchers, design of manuals, etc.;
     - monitoring, codification, data entry and correction, if appropriate, prior to analysis.
  4. *Dissemination and utilization of the results:*
     - Usually, both types of collection technique are used in the follow-up and evaluation of schemes in the following situations:

| Data required for: | Surveys | Administrative data |
|---|---|---|
| Identification of potential beneficiaries and their characteristics | x | |
| Information on the identity and number of beneficiaries | | x |
| Essential monitoring of beneficiaries' rights to benefits | | x |
| Information on expenditure involved in the payment of benefits as well as the nature and amount of other administrative costs | | x |
| Evaluation of the programme, impact analysis, cost analysis | x | |

### 4.4.5 Main collection and storage techniques within schemes

Figures 4.4. and 4.5. provide a simplified picture of the various elements of a computerized system for the storage and processing of data within contributory and non-contributory systems.

In the case of contributory systems, beneficiaries are identified when members register and on collection of contributions. This consists of administrative data. The main collection tool is the register listing membership, the payment of contributions and membership fees (depending on the system). This register may be supplemented by other supports, particularly membership cards, membership and contribution forms and in the case of health insurance certification of services provided, bills submitted by service providers and the list of services which are reimbursed.

Figure 4.4. Contributory systems

- **Recovery and control sector:** registration of employers, administration of the collection of contributions
- **Benefits sector:** registration of workers, verification of their rights and the provision of benefits
- **Financial sector/payment of benefits:** management of cash flow, capital expenditure and investments
- **Administrative sector:** management of staff, salaries, purchasing
- **IT sector:** collection and processing of data in connection with the management of the computer system
- **Statistics sector:** creation of score cards, calculation of indicators, consolidation of statistics for actuarial reviews, statistical compilations, etc.

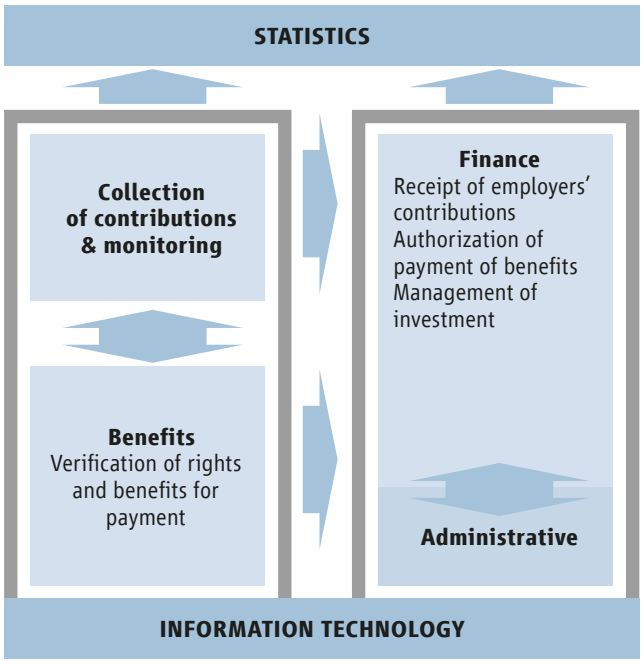
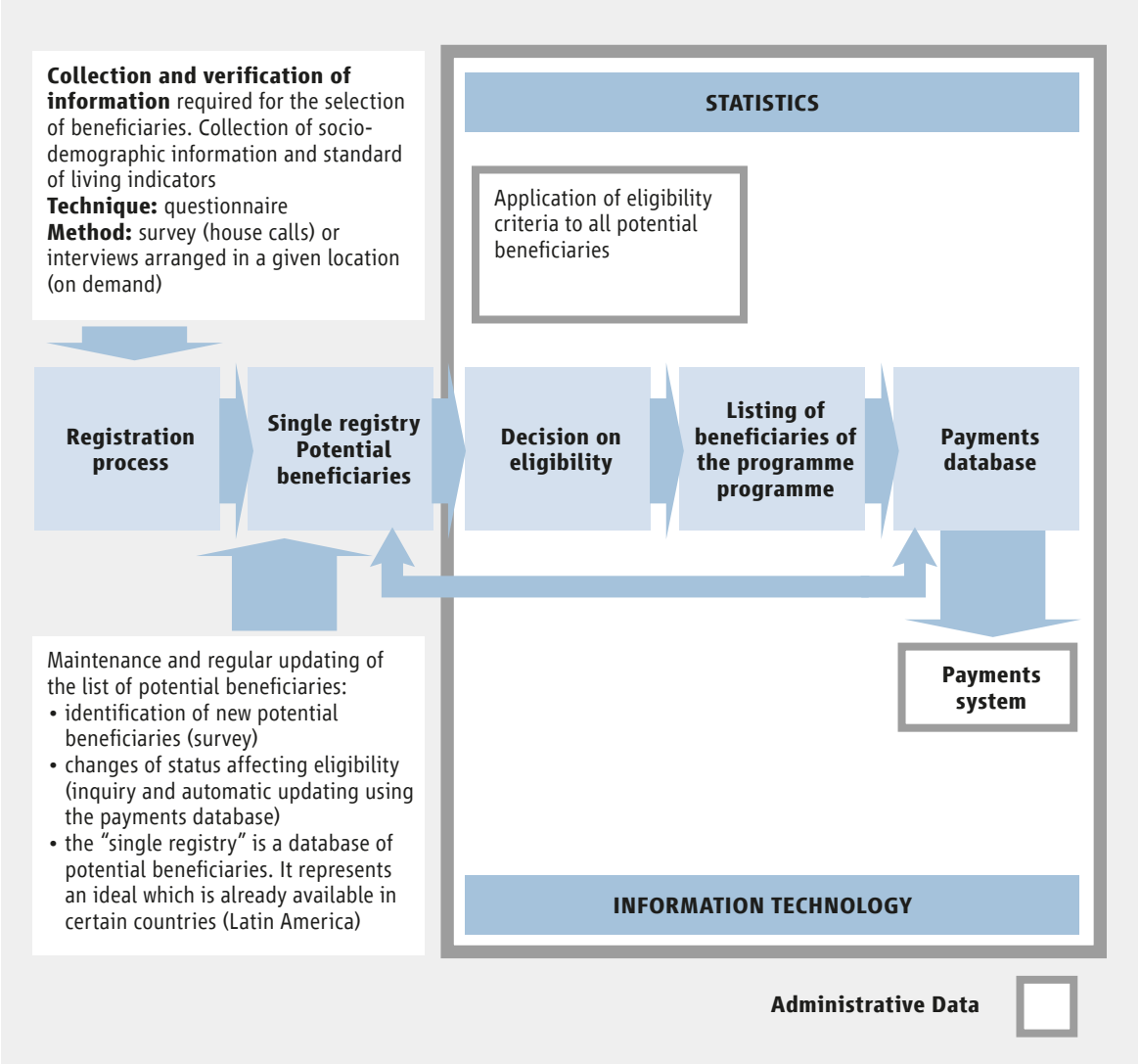


Figure 4.5. Non-contributory systems (and targeting)




In the case of non-contributory systems, the only element involved is the right to benefits. There is no contribution collection. The first step is the identification of potential beneficiaries (possibly through survey methods) followed by the listing of eligible beneficiaries.

The situation is more complex if eligibility for the programme and the payment of benefits are means-tested or involve other conditions.

- Method of identification of potential beneficiaries: identification by community, census of all households in a given geographical area, for instance through the mapping of vulnerable populations.
- Collection and verification of data: surveys (house calls) or interviews in a specified location (on demand).
- Collection tools: questionnaires to obtain all the information required to decide on the eligibility of the claimant.

### 4.4.6 Sources of data available at the national or international levels

Two main sources of data are available at both national and international levels:

1. The administrative data collected within the schemes via special questionnaires in order to enable the consolidation (often problematic) of data at the national level. This presupposes a computerized system that is up and running within each scheme for ongoing processing of data concerning beneficiaries, expenditure and receipts.
2. Information from household surveys to estimate coverage and to identify the type of benefits provided and their amount.
   - Analysis of data from existing surveys.
   - Collection of primary data through surveys.
   - Addition of a module comprising a limited number of questions concerning social security: exposure to risks, coverage provided by various formal structures in existence at the national level or by community systems, etc. This implies adaptation to each country. It also:
     - provides regular national estimates of certain elements which may indicate a need for a special independent and more detailed inquiry;
     - helps to increase awareness of the uses and necessity for statistics on social protection and particularly for consolidated statistics at the national level.
   - Independent inquiry.

Whatever the techniques chosen, they will involve prior definition of the tools (questionnaire) and, in the case of an independent inquiry, of the method of collection.

### 4.4.7 An example of data collection at the international level

The main reason for any country to show interest in data collected at the international level is to be able to compare its situation with that of other countries, particularly within the same region.

The example which follows is intended to illustrate one method of collection, storing and processing data prior to dissemination at the international level.

**Example of the collection of social security data at an international level: the Social Security Inquiry**

**Aim:** Collect data in as many countries as possible on social security expenditure, receipts, beneficiaries, the type and amount of benefits provided.
**Collection:** The inquiry collects information at two levels: nationally and at the level of the schemes.
**Collection tools:** A questionnaire comprising a defined set of questions on paper or in electronic form, with on-line records for direct integration of data in a database:

**NATIONAL LEVEL**

| • **Ministry in charge of social protection**<br>• **National institute of statistics** | • **Ministry of Finance**<br>• **National institute of statistics** |
|---|---|
| Background information<br>Population data & employment<br>**Inventory of social security schemes** | Economic & financial background information<br>Social security expenditure<br>Social security revenue |

**SCHEME LEVEL**

| **Scheme 1** | **Scheme 2** | **Scheme 3** |
|---|---|---|
| • expenditure<br>• revenue<br>• beneficiaries<br>• benefit levels | • expenditure<br>• revenue<br>• beneficiaries<br>• benefit levels | • expenditure<br>• revenue<br>• beneficiaries<br>• benefit levels |

**Collection techniques:**

- **Option 1:** Data collection at the country level:
  - Direct on-line data entry by the national institutions:
    - > **Social security schemes** for information at the scheme level
    - > Ministry of Finance: expenditure and financing on the national level, economic background information
    - > Ministry of Labour, Health: Scheme inventory, socio-economic background information
  - Hard work of identifying schemes and looking for information
- **Option 2:** Insertion of data from other sources and promotion of the use of the social security inquiry tool by other organizations
- **Option 3:** Importation of existing data at international level: ESSPROS, SOCX

## 4.5 Statistical analysis techniques

Statistical analysis techniques give meaning to the data that have been collected. The analysis enables the information to be synthesized in the form of a small number of representative statistics which can be used to:

- guide decision-making;
- identify patterns which may not be obvious at first sight;
- identify high correlations between different figures which characterize:
  - the activities of an organization or
  - the links between the organization and the outside environment.

The administrators of social security institutions are permanently seeking better ways of following up their activities and tools to facilitate good management of their operations, particularly in the financial sector. They need answers to questions such as:

- Are contributions and income from other sources sufficient to cover all expenditure on a regular basis?
- Are administrative costs well controlled in terms of the amount of benefits paid and the contributions collected?
- Is the institution getting satisfactory returns on its investments?

Those institutions which have been collecting statistical data on their schemes on a regular basis for some time are generally more confident in responding to these and similar questions concerning the conduct of their operations. Actuaries frequently complain that they are provided with little or no usable statistical data when asked to evaluate a social security scheme for the first time. This is a fundamental issue, and the administrators of social security institutions need to devote the necessary resources to it.

At the national level, the regular collection of statistical data helps to provide answers to questions facing ministries responsible for social security. These questions concern social security sectors still to be developed at the national level, population groups covered and excluded, the resources needed to finance social security, etc.

International organizations (e.g. the ISSA, the ILO, the World Bank) contribute in various ways to the development of social security measures in different countries through research projects, the organization of conferences, the coordination of financial resources, etc. These organizations also rely on statistics to assess needs and identify available resources.

The techniques described in this particular section have been restricted to the analysis of current or recent situations in order to deal with decision-taking for the short term. Other models and projections concerning the future status and financial situation of schemes will be addressed in Module 5 of this Manual.

A limited number of examples of applications of *descriptive* statistics (tables of data and/or frequency, diagrams based on primary data) and the *exploratory analysis* of data (synthetic graphs, identification of underlying structures, detection of anomalies and discrepancies) will be relied upon to illustrate the methods of analysis proposed.

Data concerning the main pension schemes (old-age, disability, survivors) for employees in the private sector in each country under consideration have been used, taken from the *ISSA Statistical Database on Social Security in Developing Countries.*

### 4.5.1 Social security schemes: Finance and administration

Figures 4.6. and 4.7. concern the pension scheme of the Barbados National Insurance Office.

These two figures indicate steady growth in income and expenditure for this scheme throughout the period 1997–2006. The level of income remains significantly higher than expenditure, and the scheme runs little risk of a deficit in the short term. The elements on the income side are all increasing harmoniously, and on the expenditure side the level of administrative costs remains stable at a low level.

Figure 4.6. Barbados National Insurance Office – Pension Scheme: Changes in the breakdown of income of the scheme (in millions of Barbados dollars)

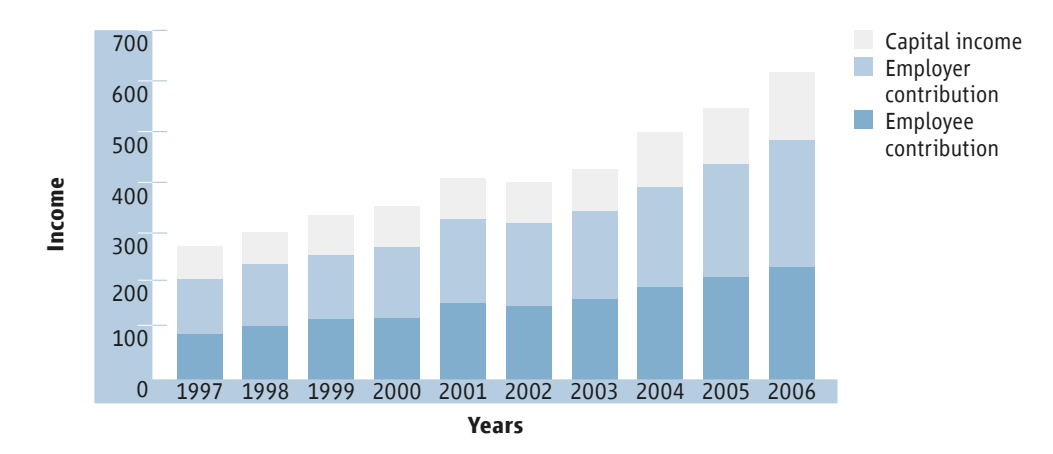
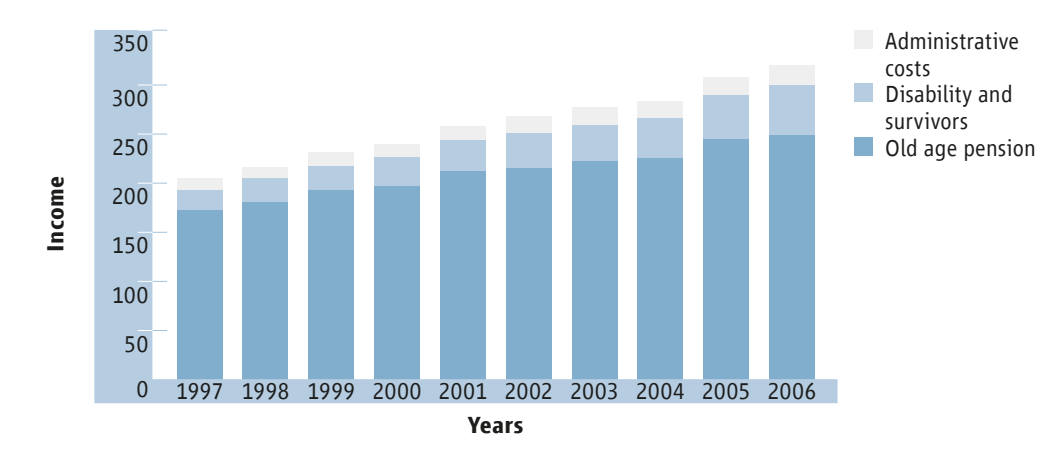



Figure 4.7. Barbados National Insurance Office – Pension Scheme: Changes in the breakdown of expenditure of the scheme (in millions of Barbados dollars)



Figures 4.8. and 4.9. illustrate two contrasting paths of development in the breakdown of expenditure in two countries in the period 1997–2005: Jordan (Social Security Corporation – SSC) and Chad (National Social Insurance Fund – CNPS).

Figure 4.8. Jordan Social Security Corporation – Pension Scheme: Expenditure of the Scheme, 1997–2005 (% of contributions)

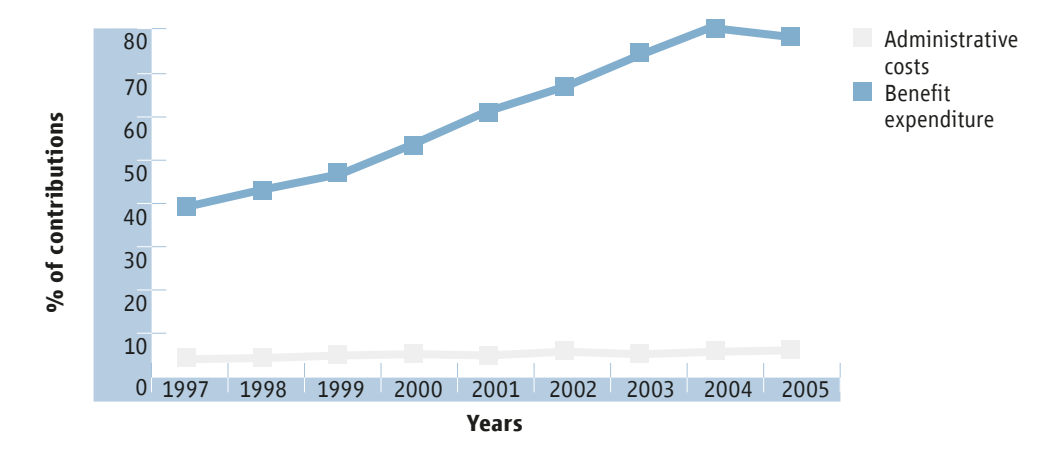
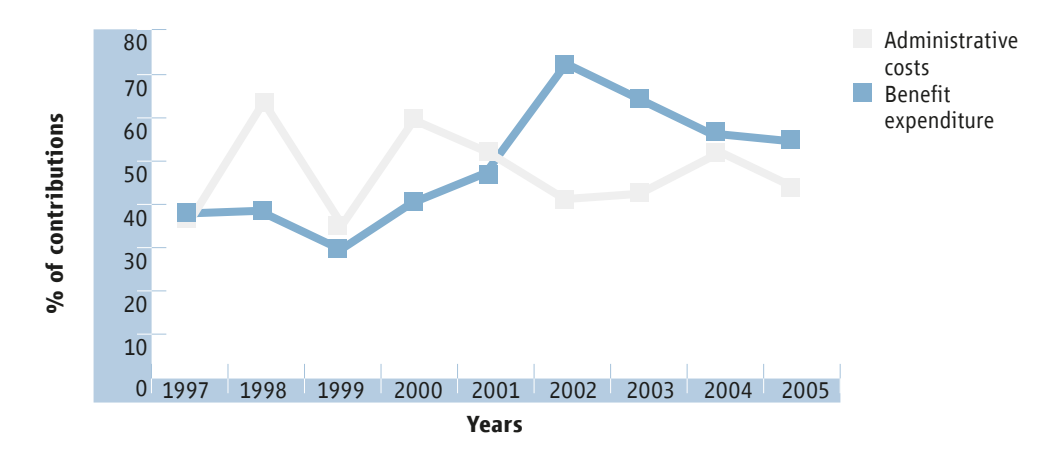



Figure 4.9. Chad National Social Insurance Fund – Pension Scheme: Expenditure of the Scheme, 1997-2005 (% of contributions)




There are a number of reasons for the differences, particularly the history of the schemes, a more efficient system for the collection of contributions and the impact of changes in the labour market or in the national economy as a whole.

### 4.5.2 Comparison of data from different countries and correlation analysis

Based on the figures below a certain number of international comparisons can be made, using the average annual pension benefit (in USD), obtained by dividing the total annual pension payments by the total number of beneficiaries (old age, disability, survivors) in the year under consideration, as the comparative variable.

Figure 4.10. compares the growth of average annual pensions in Benin, Guinea and Chad (at current rates expressed in USD, for the years 2001–2005). In Benin the average pension has increased steadily, while in Guinea and Chad it fell in the last two years, which is even more problematic for beneficiaries if inflation is taken into account.

Figure **4.10.** Average annual pensions in three African countries (USD)

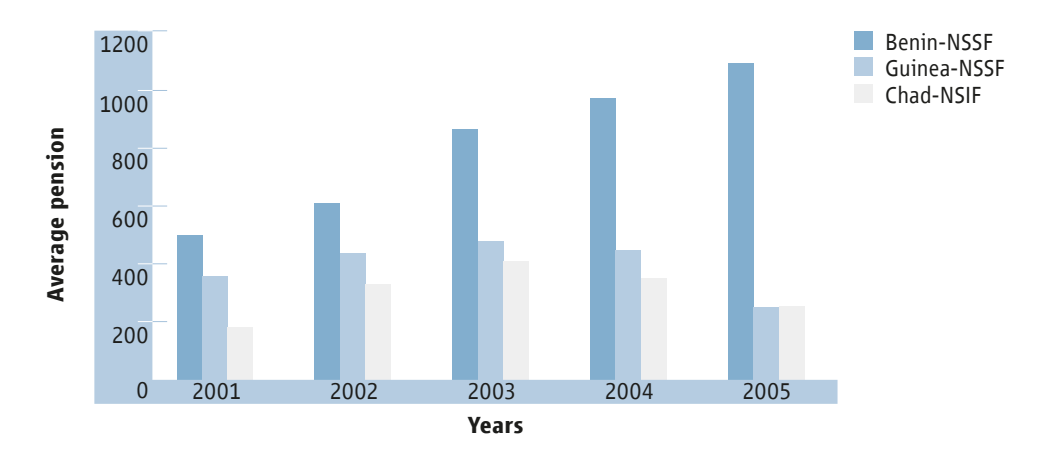



Table 4.1. provides an example of comparative data on average annual pensions in 15 countries (in USD), for the year 2003: Bahrain (2538); Barbados (3393.4); Benin (865.5); Côte d'Ivoire (715.2); Gambia (230); Ghana (523.3); Guinea (493.1); Jordan (2829.4); Morocco (1407.6); Philippines (747.5); St. Vincent and the Grenadines (1573.6); Tanzania (456.3); Chad (425.5); Togo (867.6).

Table **4.1.** "Stem and leaf": Annual average pensions in 15 countries, 2003

| | | | | | |
|---|---|---|---|---|---|
| No. of obs. 1 | *2* | 30 | | | |
| | *3* | | | | |
| No. of obs. 3 | *4* | 25.5 | 56.3 | 93.1 | |
| No. of obs. 1 | *5* | 23.3 | | | |
| | *6* | | | | |
| No. of obs. 2 | *7* | 15.2 | 47.5 | | |
| No. of obs. 2 | *8* | **65.5*** | 67.6 | | |
| | *9* | | | | |
| | *10* | | | | |
| | *11* | | | | |
| | *12* | | | | |
| | *13* | | | | |
| No. of obs. 1 | *14* | 07.6 | | | |
| No. of obs. 1 | *15* | 73.6 | | | |
| | *16* | | | | |
| | *17* | | | | |
| | *18* | | | | |
| No. of obs. 1 | *19* | 72.8 | | | |
| *No. of obs. 3* | *HIGH* | *2538* | | *2829* | *3393.39* |
| | Average | 1244 | | | |
| | Standard deviation | 960.2 | | | |
| | *Median | 865.5 | | | |

The "Stem and leaf" plot shown in Table 4.1 groups the data in ascending order and thus provides a full picture of the distribution and concentration of the data in the different quartiles. This type of diagram facilitates the rapid identification of the median and of any "discrepancies" and is fairly easy to interpret.

On the right of the column which shows the number of observations, there is a column showing the hundreds and then the corresponding observations.

Observations may be very low compared with the total or very high, as is the case with the distribution in this example, with three observations which are obviously out of line compared with the rest of the distribution (Bahrain, Jordan, Barbados). In this example, the presence of these three figures in the calculation makes the average significantly higher than the median (position indicator). Similarly, calculation of the standard deviation yields a high figure, which is one of the particularities of this data set. Without these last three figures, the standard deviation would have been far lower.

The type of classification suggested below enables the fairly rapid identification of countries or regions where the provision of support for the income of pensioners could be a top priority.

Figure 4.11. shows the correlations between average annual pensions and per capita GDP (in USD). The scatter plot shown in this figure represents combined scores (average pension: per capita GDP) for the different countries observed over a certain number of years: Bahrain (4 years); Barbados (6 years); Benin (5 years); Jordan (2 years); St. Vincent and the Grenadines (4 years); Thailand (3 years); Tanzania (5 years), Chad (5 years). Average pensions have been calculated on the basis of ISSA data, while the data concerning per capita GDP have been taken from the *World Development Indicators* published by the World Bank.

Figure **4.11.** Correlations between average annual pension and per capita GDP (Correlation coefficient = 0.789)

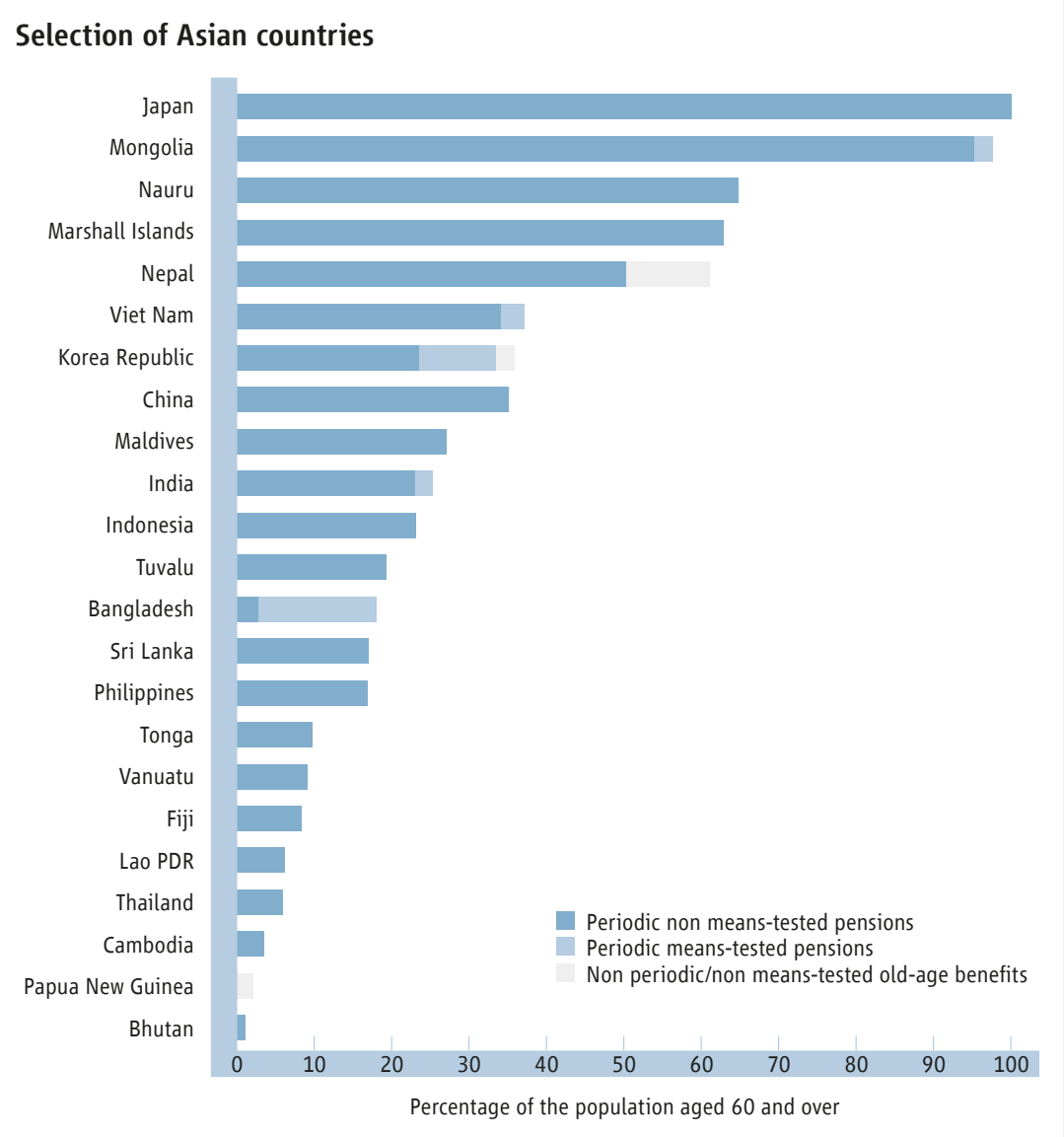

As might be expected, the situations are fairly heterogeneous. In certain countries, per capita GDP is extremely low and varies at levels well below those of the average annual pension (this is the case in most countries with a very large informal sector). In other countries, the per capita GDP is much higher than the average annual pension, for a variety of reasons.

By and large, however, there is a positive correlation between per capita GDP and average annual pension, with a correction coefficient of approximately 0.79. This supports the argument that economic growth and social protection (represented here in terms of the income provided for pensioners) should go hand in hand.

The purpose of Figure 4.12. is also to enable comparison at the international level. It provides an example of indicators concerning the old-age coverage provided in a number of Asian countries. It covers the population aged 60 and over in receipt of an old-age pension (means-tested or otherwise). The figures shown are for 2004 or 2005, depending on the country, taken from the ILO Social Security Inquiry database.

Data concerning coverage are often incomplete and more difficult to consolidate than monetary data, particularly because one person may be covered by several programmes and thus be included twice.

Figure 4.12. Proportion of the population aged 60 and over receiving an old-age pension (%)



### 4.5.3 General comments on the methods described

The aim in choosing the analytical techniques and the examples of queries provided in this section was to give an indication of the *usefulness of statistical analysis* in the field of social security in providing empirical support for decision-making in the short term. Obviously there are many major issues, and not all of them have been addressed. Issues concerning ways of improving collection and increasing the rates of coverage of national social security systems, for example, are of special importance.

In conclusion, it should be emphasized that all social security institutions need to make an effort to organize the regular collection of statistical data on the schemes they administer, because without this data it is impossible to carry out the analyses which are essential to ensure the effective management of their activities.

The importance of statistical analysis lies in the fact that it *gives meaning* to the mass of information which is collected, structuring the data in such a way as to make them meaningful and useful, enabling the identification of patterns of behaviour and relationships which up to now have gone unrecognized.

It is these statistics that enable one to say whether or not a certain theory is validated by the real facts or whether a practice carried out in a certain country could be recommended for use elsewhere. These are the factors which underline the prime importance of the use of statistical methods within social security institutions.

## REFERENCES


Bonnet, F.; Butare, T. 2007. *ILO & ISSA experience with social protection statistics and perspective* (Conference paper, Conference on the Social Protection Index for Committed Poverty Reduction, 11–12 December). Geneva, International Labour Office, International Social Security Association.

Butare, T. 2000. *Statistical data and analysis: Utility, reliability and limitations* (Conference report, Interregional Training Seminar on Actuarial Valuations, Accra, 29 May–2 June). Geneva, International Social Security Association.

Cichon, M. et al. 2004. *Financing social protection.* Oxford, Alden Group.

EPRI. 2006. *Designing and implementing social cash transfer programmes.* Cape Town, Economic Policy Research Institute.

Hagemejer, K. 2001. *Social protection expenditure and performance reviews: Methodological note on definitions, classifications and performance indicators.* Geneva, International Labour Office.

Hu, A. Forthcoming. *Performance indicators of social insurance schemes.* Geneva, International Labour Office.

ILO/STEP. 2005. *Health micro-insurance schemes: Feasibility study guide* (2 vols). Geneva, International Labour Office.
<http://www.ilo.org//public/english/protection/secsoc/downloads/publ/855p1.pdf> (accessed on 30.10.2010)

––. 2006. *Social security in low- and middle-income countries: The role of household survey data* (Issues in social protection discussion paper, No. 14). Geneva, International Labour Office – Social Security Department.

––. 2007. *Health microinsurance schemes: Monitoring and evaluation guide.* Geneva, International Labour Office.

––. 2008. *Can low-income countries afford basic social security?* (Social security policy briefing paper, No. 3). Geneva, International Labour Office – Social Security Department.

IMF. 2006. *IMF Data Quality Assessment Framework (DQAF).* Washington, DC, International Monetary Fund.

ISSA. 1997. *Innovations in information technology: A decisive factor in the further development of social security systems* (Conference report, 8th International Conference on Data Processing in Social Security, Berlin, 22–24 October). Geneva, International Social Security Association.

Noble, M; Wright, G.; Barnes, H.; Noble, S.; Ntshongwana, P.; Gutierrez-Romero, R.; McLennan, D.; Avenell, D. 2005a. *The child support grant: A sub-provincial analysis of eligibility and take up in January 2004.* Pretoria, Department of Social Development. <http://www.casasp.ox.ac.uk/reports.html> (accessed on 30.10.2010).

––. 2005b. *The child support grant: A sub-provincial analysis of eligibility and take up in January 2005*. Pretoria, Department of Social Development. <http://www.casasp.ox.ac.uk/reports.html> (accessed on 30.10.2010).

Tukey, J. W.; Wilder, J. 1977. *Exploratory data analysis*. Reading, MA, Addison-Wesley.

Wonnacot, T. H.; Wonnacot, R. J. 1995. *Statistique (économie, gestion, sciences, médecine)* (4th ed.). Paris, Economica.

MODULE 5

# Actuarial research into the future performance of social security

WARREN MCGILLIVRAY

## 5.1 What kind of questions can be investigated by actuarial methods?

Quantitative actuarial research can provide plausible answers to many questions about future national social protection schemes:

- How will future income and expenditures of the scheme develop?
- What will be the effect on the government budget?
- What financial system is appropriate for the scheme?
- Is the scheme financially sustainable at the present contribution rate and benefit levels? If not, what alternative measures to ensure sustainability should be considered?
- How does the scheme react to potential future demographic and economic developments?
- What are the current and projected future replacement rates (cash benefit/insured earnings) of the scheme?
- What are the income distribution effects of the scheme?
- What are the financing and income replacement implications of proposed reforms?

In this module, several quantitative investigations which provide answers to these and other questions are described.

## 5.2 Modelling

A modeller must *look back* in time to review the operations of a social security scheme and also *look forward* to make estimates of its future operations. For the forward-looking part of the process, a model is built that incorporates the details of the *benefit* and, if appropriate, the *contribution* and *investment* elements of the social security scheme. The model reflects the expected behaviour of factors that determine the future annual development of these components.

Actuarial research is based on the art of making optimally accurate assumptions. The assumptions incorporated in the model reflect the *judgement of the* modeller, based on the modeller's interpretation of past experience and the available evidence about the likely course of future experience. The modeller must take into account:

- long- and short-term historical data and recent experience data;
- recent amendments to the scheme;
- government policies (e.g. on inflation, migration, investments, administration);
- academic research;
- other external sources of information.

Some assumptions may gradually be changed over time from those that initially are very close to recent actual experience to others that reflect the modeller's best estimate of experience in the long-term future.

The results of a model are projections, not predictions. They do not yield a single right answer, but they should lie within a range that can be regarded as reasonable.

The major modelling assumptions can be divided into two groups:

- *demographic* assumptions dealing with changes in the insured population and beneficiaries (fertility, migration, mortality, disability, retirement rates);
- *economic* assumptions (GDP, employment, wages, prices, investment returns).

### 5.2.1 Prerequisites for modelling

Construction of a model generally requires the following data (by sex and age or age group, where appropriate).

1. *Data on current and past status of the scheme being modelled:*
   - covered population
   - insured earnings of contributors
   - contributions paid
   - benefit expenditures (by type/number/amount paid)
   - investment income
   - assets (reserve fund)
   - administration expenses

2. *Data for making assumptions:*
   - Demographic statistics
     - mortality rates
     - family composition (number of children; husband/wife age differential)
     - fertility rates
     - migration rates
     - retirement rates
     - morbidity rates
     - disability rates/recovery rates

   - Economic statistics
     - labour force participation rates
     - earnings
     - wage inflation
     - price inflation
     - investment policy/performance
     - national economic data

   The models described in this section require demographic and economic assumptions about some or all of the following factors:

3. *Principal demographic assumptions*
   Assumptions of the annual *rates of growth of the insured population* based on national population projections and assumed future *labour force participation rates*. Future *rates of unemployment* are also assumed.

   Statistics from the national statistical agency are generally used to establish *family composition*, *fertility* and *migration* assumptions. The United Nations' *World Population Prospects* provides national fertility and migration estimates.

   The *mortality* assumption should be based on the experience of the scheme. If the scheme has broad coverage, national population mortality can be assumed. In a scheme with limited coverage, if the number of deaths of persons exposed to the risk of death is statistically too low to construct a statistically valid mortality table, it is an accepted practice to modify an existing mortality table to estimate the mortality experience which the scheme is expected to experience. The mortality assumption should take into account assumed future increases in the expectation of life.

*Morbidity* refers to a physical or mental condition which results in short-term sickness and possibly long-term disability. The morbidity rate is the average number of sickness days per person per year which is equal to the frequency of occurrence of a sickness times the average duration of a sickness.

The frequency of *long-term disability* depends on the definition of disability and the application of the definition. Given the low incidence of disability and the limited disability experience of many schemes, it is not possible to produce a reliable table of rates of entry into and recovery from disability. Alternative methods of projecting numbers of disability claims can be applied. The assumed mortality table must be adjusted to produce an assumed mortality table for disabled persons.

Assumed *retirement rates* are based on the actual and expected experience of the scheme.

4. *Principal economic assumptions*
   The *salary scale* is the assumed salary evolution over the career of an insured person. It reflects productivity increases, i.e. increasing work experience, promotions, merit increments, etc.

   Based on recent past data, the *density of contributions* for contributors in a year is assumed, where

$$\textit{Density of contributions} = \frac{\textit{average number of contributions paid}}{\textit{maximum potential number of contributions}}$$

Assumed *real GDP growth* and *employment growth* are used to determine the annual real rates of productivity growth – the assumed real *rates of wage inflation*.

Assumed rates of return on investments are based on recent past returns and expected future returns. Rates of return can be calculated using the formula

$$i = 2I/(A+B-I)$$

where $i$ is the rate of return, $I$ is the investment income, and $A$ and $B$ are the invested assets (reserves) of the scheme at the beginning and the end of the year respectively.

Recent *consumer price inflation* (CPI) data on which assumptions can be based are usually available from the national statistical agency.

Generally, the following relationships among the nominal rates apply throughout the projection period for a pension scheme which follows a financial system with a level of funding.

$$\textit{Investment return} > \textit{Wage inflation} > \textit{CPI}$$

Based on scheme data on the *accrual of benefit rights*, future accruals are assumed.

Future *administration expenses* as a percentage of contribution income are assumed based on recent past experience. In schemes where coverage is expanding (i.e. the number of contributors is increasing), the administration expense ratio will normally decrease in the future.

## 5.3 Social budgets

Social budgets are based on the income and expenditure accounts of all components (branches) of a national social protection system. These components are interrelated. Social budgets provide a basis for comprehensive and coherent analysis and planning of the entire national social protection system.

The financial statements of each component of the national social protection system (e.g. health care, sickness, disability, survivors, old age, family, unemployment, housing, social assistance) provide income and expenditure data which can be used to build a comprehensive national *Social Accounting System* (SAS).

INCOME

- *Social security contributions* – by employers and protected persons; redirected social contributions
- *General government financing* – earmarked taxes; general revenue
- *Investment income*
- *Transfers from other schemes*
- *Other income* (e.g. penalties for contributions paid in arrears)

EXPENDITURE

- *Benefit expenditure* – cash benefits; benefits in kind
- *Transfers to other schemes*
- *Other expenditure* (e.g. contribution refunds, loan repayments)
- *Administration expenditure* – staff costs; purchase/maintenance of property and equipment; other

Figure 5.1. summarizes the transactions.

Figure 5.1. A social security administration as part of the national Social Accounting System (SAS)


The national Social Accounting System is a subset of the United Nations *System of National Accounts* (SNA). The SAS gives an overview, and the current dimensions of the entire national social protection system can be assessed.

Social budgets are projections and simulations that are based on the SAS by linking income and expenditure items to assumed demographic and economic developments. A social budget indicates:

- the development of all social protection revenues and expenditures for the medium-term future under the current arrangements;
- the implications for public sector budgets;
- whether measures are needed to maintain the financing of the system as a whole;
- how alternative income and expenditure measures would affect the financing;
- the sensitivity of the social protection system to alternative scenarios of future demographic and economic development.

A social budget analysis should be undertaken before any reform of existing social protection systems is introduced. Since national social protection schemes are interrelated, a change in one social protection scheme can affect other schemes.

For example, raising the retirement age for eligibility for an old-age pension may ensure the sustainability of the old-age benefit scheme. However, raising the retirement age might increase expenditure on unemployment benefits and/or disability benefits or social assistance if people leave the labour market for reasons other than retirement.

Such a scenario, of course, presupposes the existence of ample work opportunities and a comprehensive social security safety net, which is often not the case in developing countries which instead contend with patchy social security cover and high levels of unemployment.

Whatever the context though, the overall social and cost impact of a change in any national social protection scheme should be assessed before a reform is implemented.

Social budgeting is a key instrument for social protection resource management. A social budget enables short- to long-term (2–5 and up to 50 years) financial planning, management and monitoring of the entire social sector. It supports political decision-making about national social policy and financial planning. The process of establishing a social accounting system and building social budgets is an essential element of responsible governance.

## 5.4 Actuarial models

Actuarial valuations are tools for financial governance and planning of social security schemes. Actuarial valuations are undertaken for individual social security schemes to:

- assess the present financial status and likely future financial development of a scheme;
- assess the financial sustainability of a scheme under the present and possible alternative benefit provisions and financing arrangements;
- advise and recommend possible amendments to the scheme.

Actuarial valuations every three (or five) years are normally a legal requirement of a social security scheme. The legislation should specify the conditions for a scheme to be in actuarial equilibrium (e.g. annual income equal to annual expenditure; target funding level). In order to avoid a decision being taken without proper knowledge of the financial implications, legislation should specify that no amendment to provisions of a scheme may be adopted without taking into account an actuarial study which assesses the financial implications of the amendment.

Actuarial models can be deterministic or stochastic:

- A *deterministic* model is based on one given set of data and assumptions, and produces one set of outputs. Actuarial valuations applying deterministic models generally include sensitivity tests of alternative demographic and economic scenarios.
- In a *stochastic* model a given phenomenon is expressed in terms of probabilities. The output of a stochastic model is a range of possible results. Each result is associated with a probability of occurrence. The complexity of stochastic modelling means that it is usually done for specific studies.

## Box 5.1. Panama social budget

Social budget projections were designed to provide a realistic view of Panama's possible socio-economic development from 2000 to 2050.

- It is assumed that over the next 50 years the population will increase by about 50 per cent.
- GDP assumptions were developed on the basis of per capita GDP growth over the past 50 years, which was in the order of 2.3 per cent per annum. Over the projection period, assuming higher future labour productivity growth, it is assumed that the annual long-term average per capita GDP growth will be around 2.7 per cent.
- The total number of employed persons and the number of social security contributors are projected to nearly double between 2000 and 2050.
- Women's participation in the labour market will increase significantly, and it is assumed that by the end of the projection period almost no differences will exist between male and female employment levels.
- Unemployment is assumed to decline from double-digit rates to almost full employment after the mid-point of the projection period.

In 2000, total revenue of the Panama social protection sector was estimated to be around 1,921 million balboas (19.2 per cent of GDP), and total expenditure was estimated at 1,806 million balboas (18.1 per cent of GDP). Income had exceeded expenditures throughout the 1990s. In all years, contributions and tax revenue each accounted for almost equal shares of around 45 per cent of the financing of Panama's social expenditure. Capital income accounted for about 10 per cent of total revenue. Overall revenue is expected to decrease from the 2000 level to 17.1 per cent of GDP in 2050.

Social spending on old-age benefits absorbed 6.1 per cent of GDP in 2000, followed by health with 5.6 per cent and education with 4.5 per cent. Housing plays a minor role in Panama's social budget.

Over the projection period, the development of social spending as a percentage of GDP is influenced by the following trends:

- the declining relative number of children in the population will reduce the percentage of GDP spent on education;
- spending on health as a percentage of GDP is expected to rise steadily as per capita spending on health increases due to ageing of the population;
- spending on old-age, disability and survivors' benefits is expected to increase from 6.1 per cent in 2000 to more than 10 per cent of GDP in 2050;
- housing spending will remain at relatively low levels.

Given the current benefit levels, contribution and tax rates and assumed demographic and labour market developments, Panama's social protection system is expected to incur a long-term structural deficit. The deficit is expected to be 1 to 2 per cent of GDP until the late 2020s, and to rise to over 5 per cent of GDP in 2050.

*Source: Scholz et al. (2000: Box 7.3).*

### 5.4.1 Defined benefit schemes

Actuarial valuations detect emerging financial problems and indicate the need for measures to maintain a scheme's financial stability. They apply demographic and financial assumptions and systems of financing which depend on the type of social security benefit:

- *short-term benefits* – medical care, sickness and maternity cash benefits, family allowances, unemployment benefits, lump-sum grants;
- *long-term benefits* – old-age, invalidity and survivors' pensions.

Employment injury schemes generally include both long- and short-term benefits.

Short-term benefits are generally payable for less than one year. Over a period of years, the annual frequencies and average durations of these benefits are relatively stable. Hence, they have an *annual expenditure pattern* which is relatively stable as a proportion of total annual insured earnings or as an average amount per insured person.

For short-term benefits, the *pay-as-you-go* (PAYG) financial system where contributions in a period (e.g. a year) are set at a level to meet expenditures in the period is generally applied. A *contingency reserve* is set up in order to maintain stable contribution rates.

Long-term public pension scheme benefits have a future expenditure pattern quite different from short-term benefits since in general:

- each year a new group of insured persons (or their dependants) qualifies for pensions;
- pension benefits generally increase with the period of service of the insured person at the time the benefit becomes payable;
- the average new annual pension increases when the pensions are related to insured persons' earnings at or near the time they qualify for pensions;
- increased life expectancy mean pensions will be payable for increasingly long periods;
- pensions in payment are normally increased to take inflation into account.

The *PAYG curve* in Figure 5.2. illustrates the development of expenditures of a public pension scheme as a percentage of insured earnings.

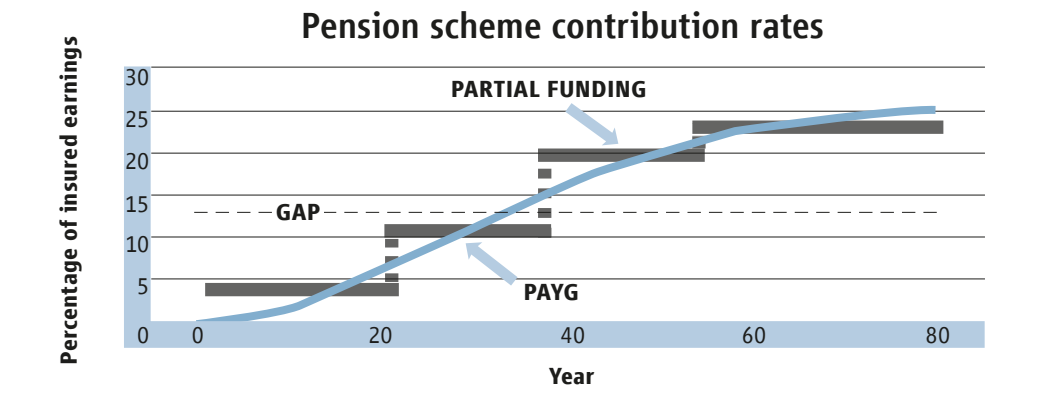
An actuarial valuation of a public pension scheme focuses on the long-term viability of the scheme. Actuarial projections (not predictions) are made for many years (100 in some cases) to assess the emerging cost of the scheme. Long-term projections are necessary to ascertain how the pension system reacts to assumed changes in economic and demographic conditions (e.g. population ageing).

Alternative financial systems are examined in an actuarial valuation. The *General Average Premium* (GAP in Figure 5.2.) is a theoretical financial system where a level contribution rate applies indefinitely.

*Partial funding systems* require periodic increases in the contribution rate (see Figure 5.2). There are many possible partial funding financial systems, for example:

- *Reserve ratio* – requires that ratio of … (reserve at end of a year) / (scheme expenditures in the year) … exceeds a certain level
- *Scaled premium* – requires that the reserve not decrease at any time.

Figure 5.2. Earnings and contributions in a defined benefit scheme




Box 5.2. Barbados public pension scheme reform

The tenth actuarial review of the *Barbados National Insurance Scheme (NIS)* was undertaken as of 31 December 1999. The review presented demographic and financial projections until 2060.

Like any social security scheme, the financial experience of the NIS is closely linked to the performance of the economy. The ability of the NIS to pay benefits is dependent not only on the contribution and benefit provisions and changing demographics, but also on the labour market and general economic activity. Continued low birth rates, net outward migration and overall population ageing can lead to shortages of skilled labour.

Key projection results in the actuarial review of the pension scheme included:
- total expenditure will exceed contribution income in 2011;
- total expenditure will exceed contribution and investment income in 2018;
- the reserves will be exhausted in 2028;
- after 2040, the pay-as-you-go contribution rate will be relatively stable at just over 30 per cent;
- a level contribution rate of 21.2 per cent would be required to meet the cost of benefits to be paid between 2000 and 2060;
- the number of contributors per pensioner will fall from 3.4 in 2000 to 1.8 in 2030 and to 1.5 in 2060.

The actuarial review included a number of recommendations and observed that the projected evolution of the scheme's income, expenditure and reserves is not unlike that of similar social security schemes.

As the NIS approaches maturity further increases in the contribution rate and/or benefit reforms are inevitable. Sustained economic and productivity growth, coupled with well-designed contribution and benefit reforms, would serve to enhance the ability of the scheme to meet its obligations to future generations.

Taking into account the actuarial review, in order to ensure the sustainability of the scheme the NIS undertook extensive consultations with stakeholders, and reforms to the pension scheme were introduced in January 2003. The principal reforms included:
- raising the contribution rate for the pension scheme by 1 per cent of insured earnings each year for four years;
- gradually raising the retirement age to 67 in 2018;
- allowing flexible retirement between ages 60 and 70 with a 6 per cent reduction in the retirement pension for each year before retirement age and a 6 per cent increase for each year after retirement age.

*Source: ILO (2001).*

### 5.4.2 Defined contribution schemes

In defined contribution *individual accounts schemes* (mandatory retirement savings schemes), the accumulated savings in members' accounts at retirement are converted into periodic payments which are made throughout retirement – the larger the accumulation, the larger the payments will be.

Members bear the *risk* that their account balances at retirement will be insufficient to provide them and their dependants with adequate retirement incomes. Consequently, projections of the amounts which will be accumulated during a member's working years are of great interest.

The estimated amount which is accumulated depends on the contributions (net of expenses) to the scheme and assumptions throughout the contribution (accumulation) period about wage growth and interest rates on members' savings.

Projections must be made over a potential *contributory period* of 40 years or more. As well as average wage growth, assumptions must be made about the density of contributions (the ratio of members' average number of contributions paid to the maximum potential number of contributions in a year).

The average accumulated balance based on these assumptions may be appropriate for a group of members, but within any group there will be great variation among individual members, with some members having higher accumulations and some lower. In addition, some members will delay entry into the scheme or have prolonged non-contributory periods (e.g. due to unemployment or sickness) which are not taken into account in the density of contributions. An individual's accumulated balance also depends on the timing of any non-contributory periods. The operation of compound interest means that contributions missed at younger ages are normally more important than those missed nearer retirement.

*Defined contribution individual account schemes* are expected to produce higher investment returns than are typically earned on the reserves of public pension schemes. Alternative interest rate scenarios should be assumed.

Clearly, multiple estimates of individual account balances at retirement can be constructed. There are no generally accepted standards regarding the assumptions which must be made. Estimates are sometimes made which serve the particular interests of the estimator.

While a projection of the accumulated balance at retirement based on assumed averages may be sufficiently robust for a group, it is unlikely to apply to any individual member of the group. Rather, in this situation a stochastic model would show a distribution of accumulated balances at retirement with a probability of occurrence associated with each balance.

## 5.5 Implicit debt

Governments are the ultimate guarantors of public pension schemes. Calculating *implicit pension debt* is an attempt to quantify the potential public pension debt of the government. Implicit pension debt is a prospective concept, the result of summing expected future deficits. Implicit social security pension debt is defined in two ways:

*Implicit social security pension debt =*

**Closed group:**

*the present value of all future benefits to present pensioners*
*+ all accrued rights of current insured persons*
*– the amount of the initial reserve of the pension scheme*
*= Implicit social security pension debt*

or

**Open group:**

*the present value of all future benefits to present pensioners*
*+ the present value of all future benefits to future pensioners*
*– the amount of the initial reserve of the scheme*
*– the present value of all expected future contributions of present and future insured persons at a constant initial contribution rate*
*= Implicit social security pension debt*

Present value at a specific time means the value of future payments discounted at an assumed rate of interest to that time. The first definition follows a private insurance concept, while the second follows a public finance approach.

Under the first definition, the calculated amount is the *termination reserve* – the reserve that would be needed in order to discharge all financial obligations to present pensioners and present insured persons with accrued rights according to the present rules of the scheme. This "full funding" reserve level is the amount which would be required to terminate the public pension scheme and honour all scheme commitments at that time.

Since public pension schemes are secured by *intergenerational societal commitments and contracts*, not by financial resources, this level of funding is unnecessary. Implicit debt calculated in this manner may be useful for some intergenerational accounting, but it has little relevance as an indicator for the overall financial status of a social security pension scheme.

For a contributory scheme, the second definition assumes that the intergenerational social contract will be honoured by present and future generations of pensioners and contributors, and that future contributors will pay their contributions at the currently legislated rates to finance the scheme. The calculated amount is the difference between expected future expenditures and revenues.

Implicit pension debt is not a sovereign debt of the government. It is a *potential* debt which indicates the *financial risk for the government*. Provided sound governance is practised, and contribution rates and/or benefit levels are adjusted in the future, the implicit debt will not occur.

A large implicit pension debt under the first definition is sometimes cited as a reason to reform the public pension system. This is true only if there is also a large implicit debt under the second definition where contribution rates and/or benefit levels are subject to future adjustments.

## 5.6 Generational accounting

Generational accounts show the present value of projected net incomes and net payments for *age groups (cohorts)* of the population over the lifetime of each cohort. For example, over the entire lifetime of a population cohort aged 20 to 30

*the present value of all income transfers and public services received*
*– the present value of all taxes and social security contributions paid*

shows whether the cohort aged 20 to 30 *profits* from the transfers and public services received or *bears a burden* to finance them, i.e. whether it has a positive or negative *lifetime revenue*.

Generational accounting can reveal the relative burdens of different cohorts or generations of the population and can be used to assess to what extent public finances are in an equitable long-term intergenerational balance. It is a tool for assessing the financial sustainability of the public sector.

Generational accounting can be applied to the national public and private (i.e. occupational and personal) pension system alone (i.e. excluding other taxes and transfers). For a public pension scheme, the objective of intergenerational equity is for all generations to pay roughly the same share of their disposable incomes during their lifetimes and have the same income replacement rates during retirement.

The intergenerational financial impact of population ageing can be analysed by applying generational accounting. The total cost of support for retired persons (the retirement burden) is the ratio of the consumption of retired persons to GDP. If a generation regards the burden as "non-sustainable", this means that it considers its share to be too high.

Sustainability in this context is a political concept, not an economic one. It means that when budget allocation decisions are taken there may be no political agreement to provide the required share of resources for the pension system to finance the consumption of the elderly.

The results of generational accounting should be regarded cautiously. They are extremely *sensitive to the interest rates* assumed to discount net incomes and net payments. Gradual expansion of coverage of a public pension scheme distorts the results of generational accounting. Generational accounting does not take into account private income and wealth transfers.

In new public pension schemes, the family support the initial generation provided to its elders before the scheme started is not taken into account; hence the initial cohorts generally receive more than the generation just beginning its working career.

Generational accounts can be used to estimate internal rates of return, i.e. the interest rate earned on pension contributions/taxes.

## 5.7 Micro simulation

Micro-simulation models deal with micro-relations between individuals or households. They are based on comprehensive data sets for the population, including, for example, information on each individual's age, sex, education, occupation, living status (single, married, household), number of children, employment status, residence, gross/net income, wealth, etc.

### Box 5.3. Developing a micro-simulation model of the South African tax and benefit system (SAMOD)

This project is in progress and is being undertaken by CASASP* for the Department of Social Development (DSD) of the South African national Government.**

The aim of the project is to develop a micro-simulation model for South Africa. The ***static*** model takes data on individuals and calculates the entitlements of individuals and households to social benefits and also the household's tax liability.

By aggregating these data to form a representative picture of the ***whole population***, it is possible to model the effect that different policy reforms would have both on national revenue and expenditure and on individual household budgets and thus the impact on poverty and inequality.

EUROMOD, a micro-simulation model – and working platform – developed for the EU Member States, has been used as the "base model" for SAMOD. Data have been drawn from the South African Income and Expenditure Survey for 2000. The model is now being refined, and options for updating the model are being explored, e.g. the 2006 Income and Expenditure Survey and the 2008 National Income Dynamics Survey (both forthcoming).

The model is being developed by CASASP in collaboration with DSD so that existing and possible future policy options for the social security system in South Africa can be explored.

Following on from the Taylor Commission in 2000, the South African Government is currently reassessing ways in which the social security system could be made more ***comprehensive*** (e.g. there are currently no benefits for healthy, unemployed people of working age who have not contributed to the Unemployment Insurance Fund).

It is intended that this model will be used to explore different policy options and to inform policy decisions within government. Members of the DSD team will be fully trained so that they can update the model themselves beyond the life of the project, to incorporate new or speculative policies and as new survey data become available.

* CASASP = Centre for the Analysis of South African Social Policy at the University of Oxford. CASASP has conducted the project with support from the University of Essex and the University of Cape Town. Publications forthcoming. Further details can be obtained from: Professor Michael Noble Michael.noble@socres.ox.ac.uk

** The project is funded by the UK Department for International Development (DFID).

Micro-simulation models are particularly useful for ascertaining the distribution effects of possible changes in social transfer schemes, in particular social assistance. This is sometimes undertaken by projecting over time, but at other times can be used to look at the "next day" impact of introducing a hypothetical new policy.

Micro-simulation models seek to describe the *dynamic experience* of individuals in the data set. The model starts with the data file on individuals at the base time ($t = 0$). By applying assumed transition probabilities (e.g. marriage, parenthood, death) and using a random variables processor, the data file on individuals at time $t = 0$ is updated to the end of the year (time $t = 1$), and so on for subsequent years.

At each stage, the information on each individual in the data file can be checked to determine whether he/she remains or becomes eligible for a benefit. For example, under the legislation applicable at the time, after applying the relevant assumptions based on individual or household income, family size/composition etc., a social assistance benefit may be payable.

Micro simulation makes it possible to model the experience of a group of individuals throughout their *life cycles* – how their experiences affect the financing of social protection programmes and how changes in legislation would affect the individuals' incomes and social protection finances.

Micro-simulation can improve estimates of social expenditure and individuals' incomes, as well as aggregate social outcomes in the branches of social protection.

## REFERENCES

Cichon, M.; Newbrander, W.; Yamabana, H.; Weber, A.; Normand, C.; Dror, D.; Preker, A. 1999. *Modelling in health care finance* (Quantitative methods in social protection series). Geneva, International Labour Office, International Social Security Association.

––; Scholz, W.; van de Meerendonk, A.; Hagemejer, K.; Bertranou, F.; Plamondon, P. 2004. *Financing social protection* (Quantitative methods in social protection series). Geneva, International Labour Office, International Social Security Association.

ESA. 2007. *World Population Prospects: The 2006 revision*. New York, NY, United Nations Department of Economic and Social Affairs – Population Division.

Gillion, C.; Turner, J.; Bailey, C.; Latulippe, D. 2000. *Social security pensions: Development and reform.* Geneva, International Labour Office.

ILO. 2001. *Barbados: Tenth actuarial review of the National Insurance Scheme as of 31 December 1999* (ILO/TF/Barbados/R.11). Geneva, International Labour Office – International Financial and Actuarial Service.

Plamondon, P.; Drouin, A.; Binet, G.; Cichon, M.; McGillivray, W.; Bedard, M.; Perez Montas, H. 2002. *Actuarial practice in social security* (Quantitative methods in social protection series). Geneva, International Labour Office, International Social Security Association.

Scholz, W.; Cichon, M.; Hagemejer, K. 2000. *Social budgeting* (Quantitative methods in social protection series). Geneva, International Labour Office, International Social Security Association.

Thompson, L. 1998. *Older & wiser: The economics of public pensions.* Washington, DC, Urban Institute Press.

MODULE 6

# Qualitative research

TIMO VOIPIO

## 6.1 Understanding how people think, feel, hope, believe and understand

The decision-makers of all organizations need to take into account the wants and needs of their customers, shareholders, members and contributors – as well as their staff. But people are complex. Their motivations and behaviour are multi-faceted and often downright contradictory. If we wish to achieve *a true understanding* of people we need research approaches and methods that can analyse and interpret such complexity in a meaningful, valid and, above all, *useful* say. This is the purpose and unique contribution of *qualitative research* (Imms, 2000).

Qualitative research is defined by many by what it is not: It is not quantitative research. It does not enable us to know the developments and trends of social security in terms of percentages or "hard facts". Qualitative research is, however, a uniquely useful method of finding out about people, *what they think, feel, hope, believe and understand*. It enables us to explore beneath the surface and to consider why people do what they do, think how they think and behave the way they behave (Marks, 2000).

Ideally, qualitative and quantitative approaches should be used side by side, in combinations adapted to the context.

In many academic circles qualitative methods of research have traditionally not been considered to be as "rigorous" as quantitative methods. In the competitive business world companies can, however, seldom afford such purism: If qualitative research methods enable companies to better predict customer behaviour – what products and services they may wish to buy, for instance – wise CEOs have good reason to take qualitative research seriously.

The same applies to social security administrations. For them it is important – but not enough – to know their coverage statistics, contribution rates and demographic trends in the population. To be able to improve the performance of their schemes social security managers will also need to know *how* people really perceive their risks and the security provided by the social security scheme. Every CEO would want to know *why* some members are deciding not to pay their contributions, and *how* things should be changed to enhance compliance and to attract new contributing members.

Qualitative methods are powerful diagnostic tools in the investigation of the *why* and *how* of human behaviour and decision-making. They can help social security administrations to understand and explain *why* people do, think and feel as they do, and not simply to state *what* they do, think and feel.

The dominant concepts and theories of qualitative research originate mainly from the social and behavioural sciences: sociology, anthropology, psychology and geography. Qualitative methods are also used as key elements in methodological approaches such as action research or actor-network theory.

Qualitative researchers use various kinds of non-directive in-depth and in-context group discussions and in-depth interviews with individuals, but can also include observation. Good qualitative research requires skills that are not the same as the skills needed in good quantitative research. The best qualitative researchers are skilful and experienced facilitators of such dialogues and interviews. Interviewing is also very different from traditional survey research in that there is no questionnaire or standard format for the interview. Instead, the researcher will prepare a topic guide, based on open-ended questions and techniques which will enable her or him to explore the relevant issues in detail. Today, sessions may be conducted in person, by telephone, via video conferencing and via the Internet.

The qualitative approach has several important functions:

- It allows respondents (those being interviewed) to talk in a way that is comfortable and meaningful to themselves, in "real" language, not that imposed by a questionnaire;
- It allows the informants to set the agenda themselves. Whilst it is important to measure responses or attitudes, in the early stages of research it is extremely useful to allow respondents to tell you what matters to them, their own starting points, their premises and their underlying beliefs;
- It encourages respondents to talk about things which they normally choose not to discuss (e.g. feelings) and about potentially important explanatory factors which the questionnaire designers would never have been able to think about;
- It enables the skilled researcher to watch and take note of the respondent as a whole – language, facial expression, body language and so forth – and, from this, to gain a much better understanding than via the use of a questionnaire alone;
- If conducted in groups the techniques can be very creative, enabling people to collectively develop and refine ideas which might be difficult individually.

Finally, qualitative analysis is very different from the analysis associated with a quantified survey. The researcher reviews all the "data" (what people said, how they responded, their body language, the silences, etc.) and comes to a view about what it all means.

An important facet of qualitative research is how findings are presented as discursive, non-statistical description. At the practical level, many research clients find this more accessible and understandable. In addition, it is typical for qualitative, participatory researchers also to engage the research client (e.g. the management of a social security administration) in the research process, and this means that the presentation of findings forms an interrogatory "triangulated discourse", with the researcher linking both the research client and respondents.

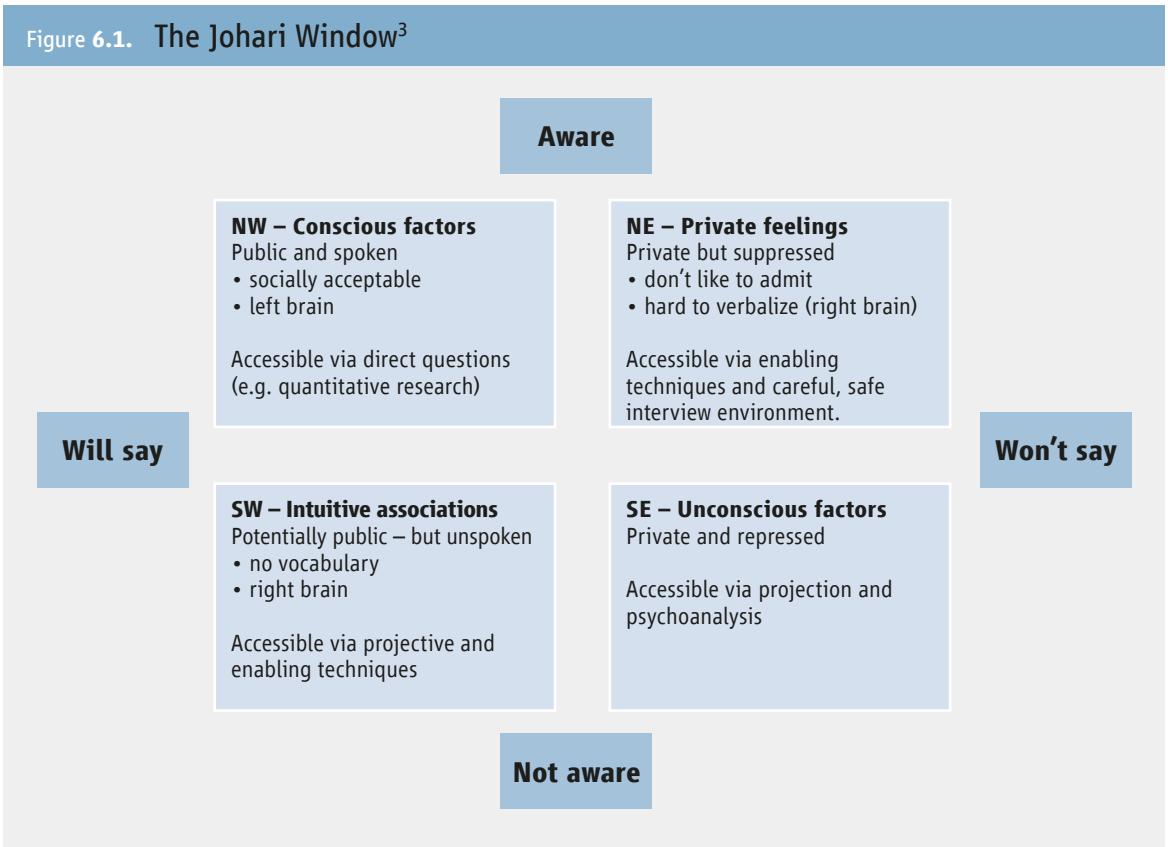
## 6.2 Some basic concepts of qualitative research

### 6.2.1 Interpersonal communication

To understand people fully – their motivations, wants and needs – we need research methods and techniques that can go beyond the conscious issues people are willing and able to express in response to simple, direct questions.

One of the ways of considering this is embodied in the *Johari Window* (Figure 6.1.) – a conceptual model for describing, evaluating and predicting aspects of interpersonal communication (e.g. a research interview). The axes of this model distinguish between those factors which people are or are not aware of about themselves, and the feelings they will and will not express.

Figure 6.1. The Johari Window[3]




NW – conscious factors:

The top left-hand ("North-West") quadrant *"Aware + Will say"* is the domain of conventional survey research. The dilemma for research is that people will tend to express views that are:

1. Socially acceptable: These include views they do not share but feel are "the right thing to say", e.g. *"We only watch nature documentaries on TV."* Moreover, it excludes things they mean but believe are unacceptable to say, e.g. *"I chose this car because I loved the statement it makes about me."*
2. "Left-brain" mental processes: Seminal work by the Nobel-psychologist Roger W. Sperry in 1968 demonstrated that the left and right hemispheres of the brain process different sorts of thoughts. The left hemisphere involves *verbal* processes – replying to questionnaires is of course predominantly a verbal activity. This gives an inherent bias in questionnaire surveys to the left-brain mental processes, including left-brain analytic processes of *rationality*, *logic* and *deduction*. In contrast, right-brain mental processes exhibit two important and distinct characteristics:
   a. They are *non-verbal*. This means that we often do not have the words to express right-brain thoughts however powerful and influential these thoughts may be;
   b. They include *feelings*, *emotions* and *intuitive associations*. Thus exploration of such factors means developing techniques to elicit right-brain, non-verbal mental processes.

While the first quadrant is of course a wholly legitimate area of inquiry for research, the importance of the Johari Window is that it maps out important dimensions *beyond* the *"Aware + will say"* quadrant. Unconscious – private and repressed – factors are not generally regarded as a legitimate (or possible) area for research, but qualitative research methods and techniques are designed to go beyond the conscious factors and to explore the other quadrants of the Johari Window: private feelings, intuitive associations and – in some cases – unconscious factors.

1. Imms (2000). The Johari Window was first designed by Luft and Ingham (1955). In their original version the four paned "windows" divided personal awareness into four different types: (1) open, (2) hidden, (3) blind and (4) unknown.

NE – private feelings:
The challenge and special skill of qualitative researchers is to create an interview environment and adopt techniques that encourage respondents to feel safe and comfortable enough to reveal *private feelings* (top-right quadrant – *"Aware + won't say"*).

SW – intuitive associations:
Qualitative researchers need to adopt interview styles and methods that allow respondents to express *intuitive associations* – things they may never have thought about before, and things they have no vocabulary to express in words (bottom-left quadrant – *"will say but not aware"*).

SE – unconscious factors:
This is the most difficult quadrant. However, if successful qualitative research processes – through a gradual progression through the three other quadrants – may reveal some information about the unconscious factors, too.

The key techniques in exploring these quadrants concern:
- Non-directive, open questions techniques;
- Active listening;
- Projective and enabling techniques.

## 6.3 The practicalities of qualitative research

Most qualitative research consists of group discussions or in-depth interviews facilitated by a qualitative researcher or moderator. The purpose and nature of such groups or in-depth interviews are summarized in Figure 6.2.

Figure 6.2. Purpose of group discussions and in-depth interviews in qualitative research

| | Group discussion or in-depth interview |
|---|---|
| **Purpose** | Exploration |
| **Moderator's task** | - To explore anything respondents feel is relevant to the topic at hand.<br>- To manage group dynamics (prevent domination by the noisy and elicit comments from the quiet).<br>- To encourage debate. |
| **Questioning technique** | - Non-directive questioning.<br>- Open, probing.<br>- Speculation, ideas and comments, including the totally unpredictable.<br>- Breadth and depth, or range of response. |
| **Output** | Understanding |

Working with a discussion guide of key topics and issues (not verbatim, standardized questions), the principal roles of the researcher at this stage are threefold:
1. To be aware of the research client's objectives;
2. To manage the process and dynamics of the group discussion or individual interview;
3. To get beyond the conscious public factors.

Methodologically, role number 2 is the most challenging. In group discussions it is essential to manage group dynamics, prevent domination by the noisy and elicit comments from the quiet. A variety of principles and techniques are used by moderators but these are inevitably situation- and circumstance-specific rather than prescriptive and mechanical.

## 6.4 Managing the qualitative research process

The following key elements have to be right in a successful qualitative research process: Open non-directive questioning techniques; projective and enabling techniques; observation; analysis and interpretation; and reliability assurance (Imms, 2000).

### 6.4.1 Non-directive, open questioning techniques

In qualitative research, the moderator retains the initiative regarding the course of the interview or discussion, but respondents are encouraged to freely relate their experience and reveal their attitudes and opinions as they see fit, with as little direction as possible from the interviewer.

Open questions usually begin with: What? Where? When? How? Why? Who? In contrast, closed questions typically begin Can you? Is it? Don't you think?'

This *psychotherapeutic* model of interviewing embraces three key principles:

1. "Transparency", including genuineness, authenticity and congruence.
2. Unconditional positive regard, and acceptance of interviewees in a non-judgemental way.
3. Empathetic understanding via attentive listening. This often includes restating what interviewees say as a way of clarifying its emotional significance and sensitivity to meanings that are just below the level of awareness.

Specifically, this introduces a fundamental requirement to create within the interview or group an *atmosphere* where respondents feel safe, respected and trustful. The so-called "warm-up" and "ice-breaker" exercises at the start of groups and interviews are often a profoundly important part of the process.

The model also introduces the need for *specific skills* and techniques, notably:

- The use of probing (e.g. "Why do you say that?");
- Active listening skills (including sensitivity to non-verbal communication) and summarizing and restating skills;
- A requirement to understand the nature and expression of emotional responses and alertness to the pre-conscious.

### 6.4.2 Projective and enabling techniques

Such techniques are key tools in overcoming the limitations of purely verbal medium.

- *Projection* involves getting respondents to reveal private feelings comfortably by "projecting" their own views on to another person or object; for example, at a simple level the question "Why do so many fail to pay their social security contributions regularly?" allows a respondent to fully express his or her reasons without admitting to having such problems him- or herself.
- *Enabling* techniques seek to help respondents reveal things they find hard to say (either because they are not socially comfortable or because they cannot find the right words).

Projective and enabling techniques commonly used in qualitative research include:

- Collage, picture sorts and drawing (visual imagery and associations);
- Word association;
- Sentence completion;
- Analogy and metaphor;
- Bubble drawings (thematic apperception tests);
- Storytelling;
- Guided dreams/visualization.

### 6.4.3 Observational techniques

Group discussions and in-depth interviews are powerful ways of exploring beliefs, attitudes and feelings. However, it has to be recognized that they do not accurately assess *behaviour*. Importantly, most people are not consciously aware of patterns of behaviour that are habitual, unconscious or culturally driven[2], for example:

- Do you know why we still have a headache after taking a ten-cent aspirin, but why that same headache vanishes when the aspirin costs a full dollar per piece?
- Do you know why people who have been asked to recall the Ten Commandments tend to be more honest (at least immediately afterwards) than those who haven't?
- People's description of their behaviour reflects their *self-image*, perceived or desired behaviour – not their actual behaviour.

For these reasons, observational techniques are sometimes used whereby researchers accompany respondents to stores, pubs, on journeys, etc., and observe and interview them about their *actual* behaviour patterns.

This can make a crucial difference to the *understanding* we gain – and is also important for social security administrations. Understanding the answer to the question about aspirin above, for example, has implications not only for your choice of drugs, but for one of the biggest issues facing our society: the cost and effectiveness of health insurance.

### 6.4.4 Analysis and interpretation

The "output" from qualitative research fieldwork (groups, interviews or observation) represents the "data", and this needs to be *analysed* formally as part of the project.

This is not simply a matter of "reporting what people said". Analysis and interpretation are integral to the progression of a qualitative research study, and cannot really be divorced from the fieldwork stages. Analysis and interpretation concern the continuous development, evolution, refinement (and often rejection) of hypotheses.

In process terms, analysis typically involves four distinct stages:

1. A functional content analysis – giving order and structure to the mass of input from the groups or interviews by sifting, differentiating, separating and sorting responses to different types and categories;
2. Interpretation – level-1: *What do respondents feel and mean?*
3. Interpretation – level-2: *What patterns emerge – and what do they mean?*
4. Interpretation – level-3: *What are the implications for the research client?* (Glen, 1997)

### 6.4.5 Reliability assurance

As a non-statistical exercise, based on relatively small numbers of respondents, it is important to consider the reliability of qualitative research findings.

Qualitative research is in no sense "statistically valid", but it is nonetheless valid – within the framework of a rather different concept of validity. There are five key facets to the concepts of validity in qualitative research (Imms, 2000).

1. Validity can be checked empirically, via *internal* consistency within the project – i.e. consistency of findings in different interviews/groups with similar types of respondents. This highlights the importance of ensuring that the project is large enough to enable such checks.
2. Validity can be checked empirically via *external* consistency with knowledge from other information sources.

2. Ariely (2008) shows that we are really far less rational than standard economic theory assumes – and that we are not only irrational, but ***predicatably irrational***. This is what the emerging field of ***behavioural economics*** is trying to understand.

3. Validity can be *maximized* by reliable sample selection and recruitment.
4. The nature of validity is quite different when exploring *shared* values and beliefs rather than *personal* preferences.
5. In a rather different way, some things *just are* – and, once they have been revealed, their validity is *self-evident.*

Certain topics *do not* lend themselves to reliable assessment via qualitative research methods. For example, price is a complex issue; while qualitative research is reliable in exploring price perceptions, components of value for money and psychological price points, it is less reliable in assessing price sensitivity, price elasticity and predicting propensity to buy.

## REFERENCES

Ariely, D. 2008. *Predictably irrational: The hidden forces that shape our decisions.* New York, NY, Harper Collins.

Butterfield, L. (ed.). 1997. *Excellence in Advertising.* London, Butterworth-Heinemann.

Glen, R. 1997. "Analysis and interpretation in qualitative research", in L. Butterfield (ed.), *Excellence in advertising.* London, Butterworth-Heinemann.

Imms, M. 2000. "The theory of qualitative market research", in L. Marks (ed.), *Qualitative research in context.* Henley-on-Thames, Admap.

Luft, J.; Ingham, H. 1955. "The Johari window, a graphic model of interpersonal awareness", in *Proceedings of the Western Training Laboratory in Group Development*. Los Angeles, UCLA.

Marks, L. (ed.). 2000. *Qualitative research in context.* Henley-on-Thames, Admap.

Sperry, R. W. 1968. "Hemisphere deconnection and unity in conscious awareness", in *American Psychologist*, Vol. 23.

MODULE 7

# Other forms of evidence – and how to use them?

CATHERINE BOCHEL

## 7.1 Introduction

This module will cover *inclusive approaches* to policy-making; other sources of evidence; collecting and drawing on these sources of evidence; evaluating different sources of evidence; and forward-looking policy-making.

The ideas presented here highlight the potential benefits to policy-makers of drawing upon different forms of evidence, and the discussion sets out some of the key ideas relevant to each area.

The forms of research and analysis discussed in the other methodological sections of this Guide do not cover all sources of evidence available to policy-makers. There are a number of other sources which are also important and which can help policy-makers at all stages of the policy process.

It is important to recognize that the other forms of evidence considered here draw upon both *quantitative and qualitative* approaches and methods, and that both can produce robust sources of evidence.

Whilst quantitative sources of data are useful in producing some forms of evidence, qualitative sources play an important role, including helping to produce policies that work in practice. Any policy, whether it is concerned with pensions, social security or employment, needs to be underpinned by a clear understanding of people's *motivations and attitudes* to the issues. Robust qualitative research can provide this.

A consideration of the type of evidence being used and whether it is appropriate for that purpose is therefore necessary. Each form of evidence should be judged on its own merits. Policy-makers need to distinguish not only between quantitative and qualitative forms of evidence, but also within these categories, since each particular form is likely to have strengths as well as limitations.

Qualitative forms of evidence, for example, might usefully make a distinction between *consultation* (which is valuable in obtaining the views of a particular population or group, but which may not be representative because respondents self-select themselves by choosing whether or not to participate) and more robust forms of research, such as *public opinion surveys* and *citizens' panels*.

*Inclusive* and "*joined up*" approaches to policy-making can play an important role in helping policy-makers devise more appropriate policies and achieve more effective policy outcomes.

The use of these other sources of evidence is important for policy-makers and researchers. However, there are difficulties in making judgements about the value and relevance of different sources, which is one of the challenges for evidence-based policy-making and why it is necessary to distinguish between different sources of evidence and to be clear about the contribution each can make.

It may also be possible and desirable to use more than one form of evidence, in order to produce more robust approaches and help address policy gaps and failures.

Social science rarely gives clear-cut answers. In addition, policy-makers will rarely, if ever, have *full* knowledge on which to base decisions. There is sometimes an assumption that evidence leads to policies that work and that it is possible to understand how policies will impact on outcomes, but this is not always the case. What works in one country, at one point in time, or in one policy area, may not work in another.

Also, policy decisions are rarely based purely on rational evidence: *political considerations*, *culture*, *values*, *costs* and the role and influence of *public opinion* and the *media* also play a part.

## 7.2 Inclusive approaches to policy-making

### What is an inclusive approach to policy-making?

An inclusive approach ensures that policy-makers take account of the impact of policies on different groups in society by *listening* to a wide range of voices, taking into account the views expressed.

It can involve the *participation* and/or *consultation* of individuals, groups, institutions, associations and organizations. These might be consumers, users, practitioners, stakeholders, citizens, other policy-makers, academics, politicians, "experts" and pressure or interest groups.

It is "*joined up*", emphasizing the need for good communication both within and among organizations, institutions, departments, research centres and bodies, so that when policy is developed it *takes account* of what is already happening, does not overlap, or repeat unnecessarily, and is built around *shared goals*.

### Why might social security institutions benefit from taking an inclusive approach?

Involving citizens alongside relevant groups and organizations and asking them for their views has the potential to improve both the policy-making process and potentially policy outcomes. The wide range of methods available to policy-makers to do this means that specific types of participation can be tailored to the aims and objectives of policy.

An inclusive approach potentially reduces problems with implementation; results can be used to inform policy; and outcomes are more relevant to target populations.

A wide variety of approaches to inclusive policy-making exist[1], and it is important to be clear about the robustness of the forms of evidence used. Many of these forms generally come under the term "participation". These can range from consultation, where policy-makers simply ask for and receive feedback on, for example, a pre-existing policy proposal, to more robust qualitative research through focus groups, public opinion surveys, citizens' panels and in-depth qualitative interviews. These approaches may also involve different degrees of citizen and stakeholder *empowerment*.

### 7.2.1 Participation

There are many different types of participation. The aims and objectives of participation must be clear to all those involved at the outset. If not, there is potential for confusion which may undermine the ability of the initiative to produce successful outcomes. It may also affect the extent to which participants are willing to get involved in future participatory initiatives.

*Policy-makers* need to ask the right questions. They have to be clear on:

- who it is they want to involve;
- why it is they want to involve them;
- which methods are most appropriate to fulfil the set aims and objectives;
- whether any groups are being excluded and, if so, why.

*Participants* need to know:

- why they are being asked to take part, for example, to provide a view on how a particular policy is working in practice;
- whether they are participating to represent themselves or to represent the views of a group or organization;
- in what ways their views will be canvassed – for example, through participating in an interview, taking part in a focus group or completing a written or online questionnaire;

1. See Bochel and Evans (2007).

- what they can expect in return – nothing at all, a meeting or written feedback to explain what particular policy options were selected and why;
- the limits to their participation so that they do not have unrealistic expectations of what their involvement can achieve.

### 7.2.2 Consultation

This is effectively a subset of participation. It has the potential to inform and influence decisions. Examples include written consultation exercises and questionnaires. This approach does not give participants any decision-making power and may not include any feedback to those who have taken part.

### 7.2.3 Impact assessment

This is a useful policy-making tool. It recognizes that individuals and groups will have different needs, but also that policies will impact on particular groups differently. Policies which are developed using impact assessment may take into account a whole range of factors including: disability, gender, ethnic, rural and urban dimensions, as well as the views of wider stakeholders. It has the potential to be more inclusive of "hard to reach" groups whose views are less likely to be heard in the policy process, and it also may give policy-makers a greater understanding of what the impacts on particular groups might be.

## 7.3 Other sources of evidence

There are a variety of different forms of evidence, some qualitative, some quantitative. A range of these are discussed below.

#### Public opinion

Policies frequently depend on consensus to be effective, so knowledge of what the public as a whole – or particular sections of the public – think is important. Such opinions are not necessarily evidence-based, and people are capable of holding inconsistent views. It is important for policy-makers to be aware of this. However, this does not mean that public opinion is not a valuable source of evidence, especially when used in conjunction with other sources of evidence.

#### Expert views

There are different types of "expert". These include:

- those who have spent years studying or working in a particular field, perhaps distinguished academics, for example;
- senior managers who may be "experts" by virtue of their long experience;
- non-traditional "experts" such as service users and local people who can contribute knowledge from a different perspective.

All of these types are valuable, but they will also have their limitations. It is worthwhile noting that "experts" do not always agree. This is generally a good thing, since it ensures that a wider range of view-points on particular issues are available for policy-makers and researchers to consider when devising policy, although it may also make the process more time consuming.

#### Practitioner views

Those responsible for administering a service will have valuable insights into what happens "on the ground" and in various parts of the delivery chain. Policies will only work if they are delivered as intended. This is why it is important to consult, listen to and act on what practitioners have to say at the formulation, implementation and evaluation stages. They can help in identifying potential problems which can then be fed back into the redesign of the policy process.

### Consumer views

Those who are in a *target group* for a policy measure or have direct experience of the service or policy measure may have useful perceptions on how the policy will be understood by consumers or how it works or will work in practice. Policies often have unintended consequences or fail to work as intended; consumers can often articulate this.

### Pressure and interest group views

These are groups that seek to influence policy-makers and the direction of policy (via, for example, lobbying and the media) in order to promote their own cause. They will bring a different perspective to the policy process, which is coloured by the interests of their particular group. Policy-makers can make use of this, whilst recognizing that it is a particular perspective. These groups may also undertake their own research and thus act as an information source for policy-makers.

### Business, professional, employer and trade union views

Businesses, professionals, employers and trade unions can seek to represent the interests of their members in the policy process by exerting pressure on those involved in the decision-making. They may aim to influence the type of issues on the agenda. They may, in some respects, be "experts".

## 7.3.1 Mediating factors

Policy-makers, researchers and bureaucrats need to take account of the mediating factors because they influence views and may even limit what is possible. These factors include political views, media, customs and values, costs and economic climate.

### Political views

All political parties have their own values and ideologies, as do other groups. These may or may not be fully evidence-based, but if a particular view or policy approach is important to those who are likely to be able to exercise power, it needs to be considered. Politicians also often have to take into consideration more and different realities than do researchers. Therefore, to expect a perfect link between research and policy decisions would be naive and unrealistic.

### Media

The media (television, newspapers, internet, radio, and so on) play a role in influencing a whole range of these other sources of evidence. They help set the agenda. They are not necessarily a neutral influence. The influence of the media may be affected by factors such as the intensity of media coverage of a particular issue, the existence of alternative sources of information, the extent to which an issue can be politicized along political party lines or associated with powerful groups, and the level of public interest.

### Customs and values

People have particular ways of seeing the world and of behaving. These will influence both their perceptions of and responses to a particular policy measure; thus, it is important to get beneath the factual and also the attitudinal. With globalization, some have argued that there has been a tendency for societies to grow more alike. This can be seen as convergence of policy goals, content, instruments, outcomes and style (Bennett, 1991).

### Costs

Virtually all policies have a price tag of some sort. Even if a policy is said to be "zero cost", it may mean that programmes elsewhere have to be cut to fund it. Whether that price is worth paying will depend on who is making the judgement (issues of self-interest, such as professional or organizational, may come into play here), the resources available, how high a priority the policy is deemed to be, and

whether it is judged to represent value for money. Policies grounded in robust data/knowledge have the potential to work better. Thus investing resources in research is vital.

#### Economic climate

The economic climate will affect the level of resources available to invest in research, which in turn will affect the level and quality of information available to policy-makers on which to base their decisions. It will also impact on the type of research methods to be employed (some are more time consuming and costly than others) and on the actual policies that can be afforded.

## 7.4 Collecting and drawing on these sources of evidence

This section looks at some of the different methods and approaches in gathering evidence to support policy and decision-making. The discussion here is broadly divided into consultative approaches and more robust research methods.[2]

### 7.4.1 Consultative approaches

#### Written consultation exercise

These exercises place emphasis on giving all those within a particular target population a chance to comment on an issue or policy.

- *Strengths*: Good for obtaining views on potentially complex policy proposals from relevant stakeholders and individuals.
- *Weaknesses*: Only gains the views of those who choose to respond to the exercise and may exclude the views of those groups or individuals who are "hard to reach".

#### Open/public meetings

An open invitation is extended to any member of the public to attend a meeting in order to find out about a particular issue. This may be done via newspaper adverts, local magazines, local radio and television, etc.

- *Strengths*: Demonstrates an open and transparent approach to policy-making and enables the organizers to present information and to respond to any questions.
- *Weaknesses*: Likely to only reach smaller numbers and depends who attends as to which views get listened to. Not representative in any way.

#### Citizens' juries

A group of 12–16 citizens recruited from a particular section of the population is brought together, usually for a number of days, to discuss a particular policy issue. Citizens receive evidence and are guided by trained facilitators.

- *Strengths*: Enables policy-makers to get an in-depth understanding of public perceptions.
- *Weaknesses*: Issues around selection and representation of participants may not have been thought through.

#### Workshops

An interactive session or sessions lasting up to two days is held which allows policy-makers and researchers to engage in a dialogue with citizens or stakeholders on a specific issue.

2. Sections 7.4.1, 7.4.2 and 7.4.3 are adapted from Cabinet Office (2002), pp. 44-50.

- *Strengths*: Provides an opportunity to explore an issue in depth.
- *Weaknesses*: It is not clear how participants are selected; nor is the degree of representativeness of the target population.

### Deliberative polls

Such policies are used to measure the opinions of citizens before and after they have had the opportunity to become informed – via briefings with experts, reading and discussion – about a particular issue.

- *Strengths*: Allows a larger number of people (250–600) to express an informed view on a particular topic and they are selected to participate as a representative sample of the population.
- *Weaknesses*: As people become more informed, they may become less "typical" of the population.

### Consensus conferences

A panel of between 10 and 20 people, recruited by random selection techniques, develop an understanding of a specific topic through briefing materials and through questioning experts or witnesses. They set out their views in a public session at the conference.

- *Strengths*: Allows the panel to be the key actor, deciding which issues to focus on, the selection of witnesses, and to come to its own conclusions; policy-making is therefore open to greater public scrutiny.
- *Weaknesses*: Panels are too small to be statistically representative.

### Issue forums

The forums are ongoing bodies with regular meetings which focus on a particular topic or issue.

- *Strengths*: Enables ongoing dialogue with participants, particularly with regard to the formulation of policies.
- *Weaknesses*: Involvement may mean that the views of participants become less "typical" as they become more informed.

### Working groups

These involve different groups such as social security organizations, pressure groups, welfare bodies, researchers, policy-makers, "experts" of all types, representatives from civil society and citizens in developing policy.

- *Strengths*: Develops policy by drawing on a range of "expert" sources.
- *Weaknesses*: Only the views of those who are selected to participate will be represented.

### Visioning exercises

Stakeholders take part in a structured meeting, where they develop a shared vision for the future and agree to action to achieve this.

- *Strengths*: Helps to create consensus among a range of different stakeholders.
- *Weaknesses*: Only the views of those stakeholders who have been asked to participate are represented.

## 7.4.2 Other approaches

### Impact assessments

The aim of this approach is to improve both the quality of the policy process and the quality of policy outcomes. It should take place at the stage where policy-makers are beginning to think about devising a policy proposal.

There are a variety of approaches, but most involve policy-makers working through a structured checklist in order to assess the possible impacts and outcomes of a particular policy on target groups.

Types of impact assessment include *regulatory impact assessment* (RIA), *health impact assessment* (HIA) and *environmental impact assessment* (EIA). When undertaken properly, an impact assessment can be both time consuming and costly. Because of this, there is a risk that policy-makers will only use it in a superficial way.

### Policy transfers

These transfers are the adoption of programmes or policies from one state or organization by another. This can take place internally, within a country, or externally, from other states or organizations.

Social security organizations confront many common policy problems, so it makes sense to utilize existing research and experience. This means that policy-makers do not have to start from scratch when confronted with a policy problem, but can look to other organizations or states to see how they have dealt with a similar policy problem.

It is necessary when transferring a policy from elsewhere to take a range of factors into consideration. These include: the likely impact on the target population; the level of resources available; cultural differences; the fact that policies are being transferred from and to like institutions and that a continuous process of evaluation takes place, so that new programmes or policies can be adapted if necessary.

### Lesson learning

Lesson learning means drawing lessons from the experiences of others. It is a dimension of *policy transfer*. Policy-makers can learn lessons from other organizations, institutions, states and individuals in respect of current and past policies. This might be learning what not to do by observing the mistakes of others, as well as learning what to do.

It is therefore important that, when drawing on the experiences of others, to not only talk and listen to a wide range of both traditional "experts" and non-traditional "experts", such as local people or users of services, but also to take into account issues around the transfer of policy (see *Policy transfer* above).

## 7.4.3 Research methods

Good qualitative research is robust. For example, proper *purposive sampling* will generate a full range and diversity of views. It can therefore be a valuable tool for exploring attitudes, understanding beliefs and behaviour.

### Public opinion surveys

Such surveys aim to obtain information from a representative sample of the population. There are different forms of surveys: for example, self-completion questionnaires, and questionnaires and checklists administered by an interviewer.

- *Strengths*: A relatively quick and cheap method of collecting data.
- *Weaknesses*: A tendency to produce numerical data which may not always explain the feelings and values that underpin the views expressed.

### Focus groups

These groups are led by a trained *facilitator*. They bring together a small group of people (typically 8–10) to discuss a particular issue.

- *Strengths*: They enable issues to be explored in greater depth.
- *Weaknesses*: Views obtained are based on small numbers of people and potentially raise issues around selection and representation.

### Reconvening groups

These groups are similar to focus groups except that participants are invited to convene on more than one occasion. They are given information to read, time to debate the topic under discussion with others outside the group and time to reflect and refine their views.

- *Strengths*: Enables participants to develop their thinking.
- *Weaknesses*: As with focus groups, this method can raise issues around the selection and representation of the participants.

### Citizens' panels

Citizens' panels are made up of a statistically representative sample of the population (ranging from 500 to 5,000 people). Their views are sought on a regular basis using methods such as interviews, focus groups and surveys.

- *Strengths*: Panels are statistically representative.
- *Weaknesses*: The representativeness of panels can be affected by attrition.

### In-depth qualitative interviews

These interviews are good for gaining people's opinions, attitudes and experiences. They can be semi-structured or unstructured.

- *Strengths*: Generate rich in-depth material and enable researchers to gain a fuller understanding of the participant's perspective.
- *Weaknesses*: Time consuming and generate large volumes of data, which require analysis; they may not be statistically representative.

### Action research

This type of research is a participatory practical problem-solving approach. Researchers, policy-makers, community groups and individuals are actively involved in identifying a problem, which they then work towards improving over a period of time. It is a robust approach and requires systematic planning. The methods for gathering information will depend on the nature of the information required. Monitoring and evaluation of the policy change are key aspects of this.

- *Strengths*: Can help in providing practical solutions to communities.
- *Weaknesses*: Time consuming and requires skilled management of the process.

### Case studies

These are in-depth studies which may be on a particular issue, community, group, household, individual, policy or event. They may involve in-depth interviews and observations, as well as other methods appropriate for the particular task. Case studies can be a means of identifying key issues for further investigation or may be a one-off exercise to illuminate how, for example, the implementation of a policy affects a group or organization. They are frequently associated with ethnographic approaches.

- *Strengths*: They provide rich data to illuminate the policy or issue under investigation.
- *Weaknesses*: They cannot make any claims to be representative. The information gained cannot be generalized to a wider population or area.

### Appreciative inquiries

These inquiries seek to engage with the research environment to promote change. They emphasize what is best about a situation or organization, in order to understand the process underpinning this and to build upon these positive aspects.

- *Strengths*: This method *builds upon what works*. It can be a way of contributing to community development and may involve harder-to-reach groups in the process.
- *Weaknesses*: The lack of direct concern with problems is sometimes viewed as a shortcoming.

### Delphi techniques

These are structured group interview techniques for seeking consensus about policy ideas and proposals. After a flow of information, ideas and analysis, a view is fed back for further discussion and ultimately a judgement.

- *Strengths*: This method is an iterative process which provides a good way of developing understanding as the research progresses.
- *Weaknesses*: It is dependent upon the ability of the group to come to some degree of consensus.

## 7.5 Evaluating different sources of evidence

The previous section has highlighted the range of different sources of "evidence" that can be called upon. However, evaluating the quality of such evidence and the weight that should be given to it is not without problems, with the very idea of what constitutes "evidence" subject to different interpretations.

The process of *triangulation*, using a range of different sources of evidence, may be helpful here because it can provide different perspectives which together create a fuller and potentially more robust picture. Nevertheless, there are a variety of assessments that might be made to help assist in evaluating these sources.

- Are policy-makers asking the right questions?
- How directly does the evidence gathered relate to the questions being asked?
- How appropriate are the methods used, and do they fit with the aims and objectives of the policy proposal? (The policy-makers need to be clear at the outset what the aims and objectives of the participatory exercise are, and they have to think about this when selecting methods and approaches.)
- Which groups have been included in the policy process, and have any groups been excluded or ignored?
  - Why have particular groups, individuals and organizations, as opposed to others, been included in the policy process?
  - Why is it important to listen to and to take into consideration the views of these groups?
  - How will their views help inform the policy process?
  - Have any groups been excluded from the process, and if so why is this? (The exclusion of people from participating in the policy process, for whatever reason, may have important effects and mean that the policy does not fully meet the intended aims.)

- Who has participated?
  - Have the full range of "voices" been included and listened to?
  - Does this include people from "hard-to-reach" groups?
  - Are a wide range of views being accessed?
- Have views been represented fairly/accurately?
  - How representative of the target population are the people who participate?
  - Are those who participate clear whether they are participating to represent their own views or the views of particular groups or organizations?
- How are the views/data collected used?
  - Who decides what to use and what to discard?
  - How is the evidence interpreted? (This could have consequences for the policy process.)
  - Is the selection or exclusion of evidence based on robust decision-making structures?

At a more general level, it is important to reflect upon some of the "softer" ideas:
- their potential impact on policy-making, implementation and outcomes;
- the circularity of the policy process. The sources of evidence contribute to policy which is evidence-based. This is policy which has been informed by evidence. Programmes or policies that are successful can provide evidence of "what works" in practice;
- the legitimacy of policy decisions;
- the transparency of the decision-making process.

## 7.6 A long-term view? Forward-looking policy-making

### What is forward-looking policy-making?

Forward-looking policy-making involves understanding issues which might affect populations ten, twenty or even more years into the future and the sort of timescales over which the impacts of changes to pensions or social security arrangements take effect.

Forward-looking policy-making therefore involves a consideration of decisions which will affect the lives of future generations. However, it is not simply about predicting the future. A forward-looking approach can help create a greater awareness of the context within which decisions are made. It can also help policy-makers understand possible alternative futures, help develop the capacity for change, and even encourage attempts to seek to shape the future (Bochel and Shaxson (2007).
- *Benefits*: This is strongly linked to "rational" approaches to the policy process. It arguably should be, but is not necessarily of itself, inclusive. It can help policy-makers make judgements about the assessment and management of risks.
- *Drawbacks*: While evidence can be problematic in short-term policy-making, considering what constitutes robust "evidence" in future work raises additional questions. There may also be tensions between short-term political imperatives and longer-term policy objectives, and it is necessary to be aware of and manage these.

### Techniques

Approaches used to inform forward-looking policy-making can be broadly divided into projections and extrapolations from current trends and judgements or predictions about the future.
- Projections. These can involve:
  - *Extrapolation* – using past and present data, such as the age of populations, to forecast future developments. However, the starting point, the time period, and the measures used may affect both the shape and trajectory of a forecast.

- *Modelling* – such as models of states' economies, which set out a series of assumptions based upon available knowledge and theoretical models, allowing a variety of options to be developed, tested and compared.

*Strengths*: These approaches can help policy-makers gauge levels of uncertainty and risk in given situations.
*Weaknesses*: Models are unable to take into account the fact that policy decisions continue to be affected by *values*, *political imperatives* and *bureaucratic and other factors*.

- Judgemental techniques. These can include:
  - *Scenario writing* – developing plausible descriptions of possible futures can help to understand them and the possible consequences of different choices;
  - *Delphi analysis;*
  - *Cross-impact analysis* – developing a list of events and possible outcomes which are then assessed by a panel of experts who consider what might happen and in what sequence. This can be helpful in showing how one situation might impact upon another.

*Strengths*: These techniques can have a value in helping to explore relationships or examining developments that might not be considered by more numerically based analyses.
*Weaknesses*: There is no way of estimating their likely accuracy.

In reality, no method – and in particular *no individual approach* – can guarantee an accurate picture of the future. Instead, arguably the major benefit is the *process of thinking* about the future and the skills and perspectives that can emerge from it, including the *ability to respond to uncertainty and change* and to consider how the future might be shaped (Bochel and Shaxson, 2007).

## REFERENCES

Bennett, C. J. 1991. "What is policy convergence and what causes it?", in *British Journal of Political Science*, Vol. 21, No. 2.

Bochel, C.; Evans, A. 2007. "Inclusive policy-making", in H. Bochel and S. Duncan (eds.), *Making policy in theory and practice.* Bristol, Policy Press.

Bochel, H.; Shaxson, L. "Forward-looking policy-making", in H. Bochel and S. Duncan (eds.), *Making policy in theory and practice.* Bristol, Policy Press.

Cabinet Office. 2002. *Viewfinder: A policy-maker's guide to public involvement.* London.

MODULE 8

# Evaluation research

ROBERT WALKER

## 8.1 What is evaluation?

Evaluation is the systematic determination of the merit, worth and significance of something or someone. Evaluations are used in all fields of business and public sector management.

Evaluators of social security policies and schemes can benefit from a variety of approaches and methods of policy and programme evaluation.

Good programme and policy evaluations assess the performance of programmes and policies, measure their impacts on individuals, families, communities and national development goals, and document the successes achieved, or the shortcomings.

With this evaluation information policy-makers and programme level decision-makers are able to direct limited resources to where they are most needed and most effective for the contributing members and the wider communities.

Evaluation is a methodologically diverse field. People who carry out evaluations come from many different backgrounds: sociology, social policy, psychology, economics, social work, political science and public administration. Evaluations can involve quantitative or qualitative methods of social research, or both. Case studies, survey research, experiments and model building are among the typical methods used. A selection of methods will be introduced in this section.

### 8.1.1 What is social security scheme evaluation?

Social security scheme managers and staff frequently informally assess their scheme's effectiveness:

- Are contributors benefiting from the scheme?
- Are there sufficient numbers of contributors?
- Are the compliance levels satisfactory?
- Are the strategies for the extension of the schemes working?
- Do staff have the necessary skills?
- Is the scheme financially sustainable?

These are all questions that scheme managers and staff ask and answer on a routine basis.

Social security scheme evaluation addresses these same questions, but uses a *systematic method for collecting, analysing, and using information to answer basic questions about a social security scheme* – and to ensure that those answers are supported by evidence.

This does not mean that evaluation is beyond the understanding of scheme managers and staff but it does mean that technical expertise and experience of evaluation must be called upon in conducting an evaluation.

### 8.1.2 What is social security policy evaluation?

Social security policy evaluation is *research* that answers particular kinds of questions about the workings of social security policy and policy interventions.

Social security administrations do not always have the final say on social security *policies* because policy-making is the prerogative of governments and parliament. But it is often in the interest – and also the responsibility – of researchers and managers in social security institutions to conduct policy research and policy evaluations that will help ministers and members of parliament to make good decisions about social security policy.

Having received a mandate and a task from the national policy-making organs, it is then up to the social security administrations to operationalize the national social security policies. A great amount of "internal policy-making" within the social security administrations is required for turning the political will into practical implementation of social security schemes. This internal policy-making of social security administrations must also be evaluated from time to time.

Thus, policy evaluation – as it is discussed in this Guide – can focus either on the national social security policy-making, or on the internal policy-making within social security administrations.

### 8.1.3 Why evaluate social security schemes and policies?

You should evaluate your policies and schemes because an evaluation helps you accomplish the following:

- find out what is and is not working in your schemes and policies;
- show your members and the policy-making community what your schemes do and how they benefit the current members, and for what cost;
- get new contributing members to your schemes, or new policies mandating you to extend social security coverage, by providing evidence of the effectiveness – and gaps – of the existing schemes;
- improve your staff's work with members by identifying weaknesses as well as strengths: what does and does not work in your type of schemes with your kinds of members.

### 8.1.4 Rationale for evaluation

Evaluation is both an aid to good social security policy-making and a stimulus for good governance of social security schemes. Evaluation contributes to good governance in a number of different ways:

- Choosing between alternatives
  It may allow decision-makers to choose between competing alternatives which are shown to be differentially effective (Walker and Duncan, 2007). In making such choices, it is important not only to focus on average impacts but to assess how best the policy or scheme should be targeted by considering the relative sizes of groups that exhibit the most positive effects.
- Establishing cost effectiveness
  Good evaluations include a cost-benefit analysis which indicates whether benefits exceed costs, judged from different perspectives: typically those of the administration, government, society and the contributing members.
- Informing strategic resource allocation
  Evaluation results may be used to influence strategic resource allocation decisions quite distant from particular interventions. Strong evaluation evidence may give one organization – or scheme – an advantage over another when competing for limited public sector resources.
- Improving implementation
  Evaluation, especially formative evaluation, has a focus on implementation, ensuring that the processes of delivery are optimally attuned to fulfilling policy objectives. Creatively used evaluation can foster a process of reflexivity among all those engaged in the delivery of social security schemes and benefits especially when participative modes of evaluation are employed or when staff are actively consulted and engaged in a mixed methods evaluation.
- Enhancing accountability
  Evaluation makes real the prospect of both internal and public accountability. Without a clear statement of objectives and evidence to measure outcomes against target objectives, it is impossible to assess whether:
  - policy is effective;
  - achievements match political rhetoric;
  - resources are used wisely.

- Stimulating institutional learning
  Systematic policy evaluation can stimulate institutional learning both at the level of scheme implementation and policy-making with benefits that far exceed the insights gained from evaluating particular schemes only. Indeed, the formal articulation of goals and mechanisms demanded by formal evaluation offers a rigorous model for policy-making in general.

### 8.1.5 Evaluation as a stimulus to good governance

Regularly asking whether policies and schemes are working, or would work if implemented, implies, and requires, institutions that place *evidence* at the heart of the policy- and decision-making processes.

While public policy-making is often overtly political, with policies shaped by politicking and presentation, evaluation encourages due attention to be given to facts and rational understanding alongside other considerations.

- From audit to evaluation
  Until recently, policy-making was dominated by audit. Evaluation was located at the end of the policy-making process, as follows:
  – a problem is identified;
  – a policy solution is devised;
  – the policy is implemented;
  – the outcome is monitored and evaluated.

  Evaluation was backward-looking, asking what had worked rather than what can work.

  Evaluation offers more than accountancy or an exercise in good policy husbandry, knowing what a particular social security policy has delivered. Building in a concern with the effectiveness of policy from its inception, it can:
  – improve decision-making and aid resource allocation;
  – enhance accountability;
  – foster organizational learning (Cabinet Office, 1999).

  This requires establishing a culture of evaluation within the policy process and in the regular governance of social security schemes.

- Evaluation as a process
  With evaluation conceived as a process rather than as an event, decision-makers are encouraged to:
  – Set out the desired policy outcomes and to consider what would be the scheme outputs that could most effectively, economically and sustainably lead to the achievement of those outcomes.
  – Determine what evidence is available, relevant and useful, what evaluation systems and performance targets are required and what alternative solutions might productively be considered.
  – Pay more than usual attention to the process of implementing policy, the mechanisms of scheme delivery and the significance of the frontline staff, and especially supervisory level management, in attaining outcome goals.
  – Become more sensitized to the time required to set up schemes and management systems, for procedures to develop and structures to adjust before there is any semblance of the stability necessary for meaningful measurement.
  – Become more acutely aware of the gap between aspiration and feasibility than in the past.

## 8.2 What are the questions that evaluations can answer?

There are two principal kinds of evaluation question:

1. Primary questions, which *summative evaluations* seek to answer.
2. Supplementary questions, which *formative evaluations* provide answers to.

Regardless of the kind of evaluation, all evaluations use data collected in a *systematic manner*. These data may be quantitative such as counts of contributing members of a social security scheme, or of amounts of benefits or other services received, or incidence of a specific behaviour. They also may be qualitative such as descriptions of what transpired or an expert's best judgement of the functioning of the scheme.

Successful evaluations often blend quantitative and qualitative data collection. The choice of which to use should be made with an understanding that there is usually more than one way to answer any given question.

### 8.2.1 Summative evaluation: Primary evaluation questions

The primary evaluation question asks: *"Is the policy – or scheme – working?"* Evaluation that addresses this question is often called *summative* or *impact evaluation*. (Further discussion of summative evaluation may be found in sections 8.4 and 8.5.) Answering the primary question determines the need or otherwise for policy/scheme modification or replacement.

To establish whether a policy or a social security scheme works requires specification of *objectives*; specification of a *counterfactual*, and specification of the *baseline*.

- *Social security objectives*
  Social security policies and schemes may have many different types of objectives:
  - *Endogenous or internal objectives*: Objectives of those responsible for designing the policy or benefit scheme.
  - *Exogenous or external objectives*: Objectives attached to a policy by stakeholders and others not responsible for the policy/scheme design.
  - *Overt objectives*: Internal objectives that are publicly stated.
  - *Covert objectives*: Internal objectives that are not made public.
  - *Primary objectives*: Objectives that provide sufficient cause to introduce a policy or benefit schemes where none existed.
  - *Secondary objectives*: Objectives associated with rectifying deficiencies in existing provision.
  - *Functions*: Consequences of policy outcomes, intended or otherwise.

  Policies and schemes can be evaluated against all kinds of objectives. Which objectives are chosen primarily depends on the purpose of the evaluation and the organisation which is conducting it:

  a. *Textbook evaluations*: Textbooks tend to presume that policies and schemes will be evaluated against *internal* objectives. This is likely when the evaluation of social security schemes is undertaken by social security organizations or by governments.

     Social security schemes often have multiple objectives and it is rarely possible to evaluate schemes with respect to all of these objectives or, at least, to do so with equal precision; in such circumstances, policy objectives need to be *prioritized*.

     When objectives of social security schemes are not clearly stated in public documents, they may have to be presumed. If the objectives are deliberately covert, effective evaluation is generally only possible when undertaken by organizations privy to the objectives when, to maintain the subterfuge, no published evaluation may be produced.

b. *Exogenous evaluations*: Sometimes evaluations are undertaken with respect to exogenous objectives, ones that the policy was never intended to meet. This often occurs in international comparisons when, for example, social security polices are evaluated with respect to their impact on poverty even though, in some jurisdictions, the goal of policy is to foster social cohesion rather than to alleviate poverty.

Evaluations conducted by pressure groups or political oppositions may also adopt exogenous objectives, typically ones suited to their own cause.

It is sometimes highly appropriate to evaluate policies against objectives not recognized by architects of the policy. For example, policy objectives ascribed to an existing policy may evolve faster than designed in which case it would be appropriate to ask how well the policy meets the new objectives.

It is important to recognize that policies often have *unintended consequences* and that evaluations should also address the functions of a policy as well as the endogenous objectives.

In practical terms, it is usually unavoidable that policies and schemes will only be assessed against a subset of functions.

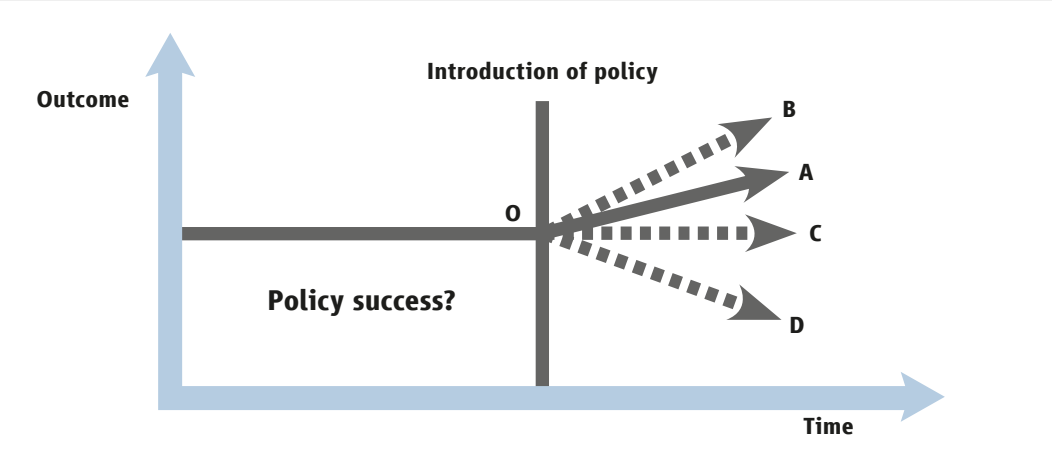
- *Counterfactual*
  Summative evaluation is inherently *comparative*.

Implicit in the primary question "does it work?" is the idea of change or difference. Does the change in the policy or in the benefit scheme matter? Is the situation different from what it would have been had the policy – or new scheme – not been introduced?

The situation that would have occurred if the policy had not been implemented is called the *counterfactual* (see the diagram below). All research designed to establish the outcome of policy involves a counterfactual.

Experimental and quasi-experimental designs described below make the counterfactual explicit; some other designs – either intentionally or through neglect – do not do so, which makes it difficult to assess the true impact of a policy.

Figure **8.1.** The importance of a counterfactual



The diagram portrays trends in an outcome measure, say income, before and after the introduction of a policy. The upward trend in income, OA, would suggest that the policy was a considerable success but not if the counterfactual, what would have happened in the absence of the policy, was OB. Likewise, the situation of no change, OC, does not necessarily mean that the policy was ineffective if incomes would have fallen, OD, had the policy not been introduced.

The challenge for the evaluator is that the counterfactual cannot be observed directly, it simply does not exist. The evaluator therefore has to engage in *"as if"* thinking: what would have happened had there been no policy. Experimental and quasi-experimental methods attempt to make this *"as if"* thinking transparent using a range of computational methods to simulate the counterfactual.

- *Baseline study – baseline data*
  Summative impact evaluation usually requires the collection of baseline data, ideally prior to the launching of the new social security policy or scheme. Baseline data should be collected – normally through a baseline study or from existing administrative data sources – for both an intervention group and a comparison group.

  A second round of data collection is needed at the beginning of the impact evaluation, for instance after the new social security scheme or policy has been in operation already for some years. Comparing outcome variables from the second round of data collection with those from baseline provides a first estimate of the impact of a policy. Comparing the differences for both the intervention and control groups provides a more accurate estimate of policy impact.

- *Impact evaluation and impact assessment*
  It is important to underline here that summative *impact evaluation* differs markedly from *impact assessment*. The latter entails a structured examination of policy proposals or social security scheme designs, or occasionally existing policies/schemes, in order to assess – mostly in advance – whether the implementation of the policy or scheme is likely to have disproportionate adverse or favourable effects on particular groups.

  For the most part, impact *assessments* (e.g. Poverty and Social Impact Assessments, PSIAs) are based on critical analysis (typically *ex-ante*) of the proposed policy principles or scheme design features with some modelling or possible scenarios.

  Impact *evaluation*, in contrast, involves direct empirical investigation (typically *ex-post*) of what actually happened when a policy was implemented.

### 8.2.2 Formative evaluation: Supplementary evaluation questions

The primary question of whether a policy or a social security scheme works is generally complemented by the supplementary ones:

1. "How does the policy/scheme work?"
2. "Why does it not work better?"

Answers to the supplementary questions can indicate the nature of the policy change required to achieve the policy objective specified. Evaluation that addresses these supplementary questions is usually called *formative evaluation*.

Supplementary questions lend themselves to qualitative research of which there are several different styles. Many large-scale evaluations include both summative and formative evaluation and hence typically combine quantitative and qualitative methods (with varying degrees of success). Indeed, evaluation studies are one of the drivers of the burgeoning interest in "mixed methods" research (JMMR, 2007; Walker, 2008).

### 8.2.3 When evaluation questions are asked

The grammatical tense in which questions are asked (past, present or future) often helps to determine the most appropriate kind of evaluation research to undertake (see Box 8.1.):

1. If questions relate to *current or past* policies or schemes, impact evaluations typically depend on the baseline data to compare the outcomes of the policy or scheme to the information available about the situation that existed before the current or past policy or scheme was launched. In situations where no proper baseline data exists, it is often difficult, though not impossible, to define a counterfactual.
2. Questions that address the *future*, and concern policy ideas that are *well formulated*, are ideal candidates for both impact and formative evaluation. It may be that:
   a. It is unclear how well they would work in practice and, hence, whether large-scale implementation would prove cost-effective.
   b. There are competing models and there is a need to choose between them.

Policies that are *not well specified* do not lend themselves to evaluation, although the decision to evaluate may trigger the detailed specification process.

**Box 8.1. Questions of evaluation**

| Time perspective | Evaluation question | Illustrative evaluation method(s) | Counterpart formative evaluation question | Illustrative formative evaluation approaches |
|---|---|---|---|---|
| Extensive past | What worked? | Meta-analysis, Systematic review | How did it work? | Systematic review |
| Past | Did the policy work? | Retrospective evaluation | How did it work/not work? | Retrospective interviews and case-studies, Participative judgement |
| Present | Is this policy working? | Monitoring (interrupted time series and natural experiments) | How is it working/not working? | Process studies, Implementation evaluation, Ethnography |
| Present to future | Is there a problem? | Basic research, policy analysis | What is the problem? | Basic research, Rapid reconnaissance |
| Close future | Can we make this policy work? | Prototypes, micro-simulation | How can we make this policy work? | Theory of change, Participative research, Action research |
| Future | Will this policy work? | Programme evaluation (impact or summative evaluation) | How would it work/not work? | Theory of change, Laboratory evaluation |
| Expansive future | What policy would work? | Prospective evaluation: micro-simulation, laboratory experimentation, gaming | How would it work? | Laboratory evaluation, Delphi consultation, Gaming |

*Source: Adapted from Walker and Duncan (2007)*

## 8.3 The management of evaluation

As with any deliberate process, evaluation needs to be managed. Management responsibility can be variously located and the tasks vary accordingly.

Quite often responsibility for the evaluation function is placed close to the policy- and decision-making process, either within government or social security organizations. Typically the day-to-day management of evaluations is assigned either to personnel with research expertise close to the top management, or delegated to maximally independent specialist evaluation sections. The practical evaluation work can be done by staff or subcontracted to universities or research agencies.

Alternatively, the locus of control may be passed to staff engaged in the process of delivery close to the interface with clients. In such cases, support and mentoring structures and procedures need to be put in place so that staff can:

- create the procedural distance to facilitate engaged but objective interpretation;
- recognize when expertise is required;
- call on expertise in judicious and timely fashion.

### 8.3.1 The process of evaluation

Evaluation begins with the decision to evaluate a social security policy or scheme and the earlier in the life of the policy, the better and more useful the evaluation is likely to be. Two pure-form models can be identified: evaluation-led policy and policy-driven evaluation. Neither model is frequently found in its pure form, but reality is usually more like the latter than the former.

1. *Evaluation-led policy*
   Evaluation-led policy entails the intention to evaluate from the inception of design work on a policy. As a consequence:
   a. Policy development work gains from the analytic rigour imposed by evaluation.
   b. The evaluation can be sensitively attuned to the objectives and institutional design of the policy.

2. *Policy-driven evaluation*

If evaluation-led policy is the ideal, policy-driven evaluation is more widespread. Under this scenario:

a. Policy design is driven by the aspirations of policy-makers and shaped by unanticipated events or crises.
b. Policies are complex with multiple, often imprecisely defined objectives that are variously tailored to different subgroups.
c. Implementation is often intricately varied.

Such policies are difficult to evaluate and require great creativity and artistry in the design of evaluations; each evaluation needs to be bespoke, designed in response to the characteristics and constraints that define the policy.

From the perspective of evaluators designing the evaluation, the earlier they can be engaged and the more briefing they can be given about the policy/scheme and the rationale for the evaluation, the better able they will be to design an evaluation that meets the needs of the organization.

### 8.3.2 The stages of evaluation

To simplify, there are generally four key stages in the evaluation process:
1. Analysis of objectives and the intended design;
2. Design of the evaluation;
3. Implementation;
4. Analysis, interpretation and dissemination.

1. *Analysis of objectives and the intended design*
   - Specify the original objectives
     The first task is to identify and clearly specify the objectives of the social security policy or scheme which is the focus of the evaluation. Then one needs to prioritize the objectives according to their importance in terms of knowing how well the entire policy – or scheme – is performing. This reflects the fact that it is rarely possible to evaluate a policy against varying objectives with equal precision.

   - Define indicators and other performance goals
     Policy objectives next need to be translated into performance goals specified in terms that can be measured or otherwise reliably assessed and verified.

     These days it is commonplace to define measurable or verifiable performance indicators for most policies, programmes and social security schemes already at their design phase. But if not, such performance indicators have to be defined by the evaluator at the early stages of the evaluation process.

     To take an example: if the objective is to reduce poverty, the performance goal might be stated as a specific percentage point change in the poverty rate.

     It should be noted, however, that income poverty is just one of the many dimensions of poverty and that even income poverty is measured differently in different countries, e.g. as the proportion of population with individual disposable incomes below the equivalent (PPP) of one, two or four US dollars a day (depending on the average level of incomes in the country) or below 50 or 60 per cent of median equivalized disposable income.

     It is important to know whether the politically defined national poverty reduction goal has been defined as lifting a maximum number of people out of income poverty, or as lifting the average incomes (which could happen even if only the incomes of the non-poor grow, and the poor remain as poor as before). In some countries more structural indicators have been defined for measuring changes in poverty, such as reductions in income inequalities (gini) or increases in the number of people engaged in decent employment.

     It is also important to state how much improvement in the goal is required for two reasons:
     - to remove the political temptation to laud any improvement, however small, post facto as a policy success; and
     - to permit the sufficiently large samples to be drawn for the purposes of the evaluation for statistical tests so as to detect changes of the order of magnitude required.

   - Re-articulate the assumed theory of change, limitations and risks
     There is great merit in re-articulating at the beginning of an evaluation the *theory of change*, as well as the *limitations* and *risks*, as they had been articulated by the original designers of the intervention. The task of the evaluators, then, is to determine:

- to what extent the real developments have met the expectations of the original planners;
- to what extent the operational context has changed in ways that the original planners had not been able to anticipate.

It is the task of the policy architects – the politicians and policy-makers – to specify openly the theory of change that has guided their planning work. They also have to specify which limitations and risks they could anticipate at the moment they designed the intervention. In doing so they will need to focus on – for example – the contribution of the following to attaining outcome goal:

- the process of implementing policy;
- the mechanisms of delivery;
- the role and significance of frontline staff;
- the role and significance of supervisory level management; and
- the circumstances, aspirations and behaviour of clients.

A clear theory of change explicitly stated before evaluation ensures that, after the evaluation, one not only knows whether the policy works but how and why. Only with a theory of change is it possible to determine in which respects the policy failed to work or which elements are crucial to its success.

2. *Design of evaluations*

Even if evaluators are not involved in policy/scheme specification, they need to be engaged in evaluation design. A number of model designs are discussed below but, while there are general design principles, there are no recipes. The art of good evaluation lies in the blend of response to both design and practical considerations

- Design considerations

  The key design considerations include:

  - Whether the policy or scheme is new or already implemented: the latter significantly narrows the types of evaluation possible;
  - Whether the policy/scheme is local or national: the latter limits the scope for defining a counterfactual;
  - Whether the political profile of the policy or scheme is low or high: while more resources are likely to be available, the latter may still be more difficult to evaluate since results will be considered sensitive and early findings may be hijacked to change the policy being evaluated;
  - Whether the policy/scheme is targeted on people or places: the latter tends to be more expensive since research has to be conducted in a number of places;
  - Whether system-wide effects are anticipated as well as individual ones: the former tends to be more expensive since research has to be conducted on entire systems (often defined as geographical areas);
  - Whether the policy/scheme aims to change circumstances or behaviour: the latter is more difficult since behavioural changes are harder to measure and generally take longer to become apparent;
  - Whether the policy/scheme marks a large change or a small one: the latter may be expected to have a smaller effect that will be difficult and expensive to detect;
  - Whether the policy/scheme has a few goals or many: the latter is likely to be more expensive to evaluate because of increased complexity;
  - Whether the policy/scheme is simple or multi-faceted, targeted on one group of interest or many: the increased complexity adds to the difficulty and expense of evaluation;
  - Whether the policy/scheme is implemented by a single agency or by multiple agencies or partnerships: the latter is more complicated to evaluate and makes it especially difficult to undertake a cost-benefit analysis.

- Practical considerations
  The evaluation designer is likely to have to react to a range of more or less reasonable demands many of which are mutually exclusive. For example, policy-makers may:
  - Hold conflicting expectations about what they want from evaluation;
  - Wish to demonstrate that a policy works, rather than determine whether or not it does;
  - Be keen to learn how to fine-tune and implement a policy rather than wait to determine whether it has the desired effect;
  - Be eager to cherry-pick components of a policy or implementation without paying much attention to the context that allows the cherries to appear ripe for picking.

  Evaluation timetables are generally tight, with pressure for quick results competing with concerns over robustness. Therefore, designs are often a compromise between the ideal and what can be achieved within a preordained policy timetable.

3. *Implementing evaluations*
   The success or otherwise of evaluations often lies in the hands of staff responsible for delivering the policy or managing the scheme. This discussion is limited to staff in charge of day-to-day management of the scheme or policy.

   Evaluators typically have to balance the needs and concerns of the staff against demands for rigorous evaluation. This is not always as straightforward as it might appear. Staff delivering services may:
   - Aspire to minimize inconvenience and maximize benefits for their clients and for themselves;
   - Not want to assign individuals whom they believe would most benefit from the intervention into a control group denied the intervention;
   - Be anxious about the prospect of their own work being appraised; and
   - Have limited understanding of evaluation research.

   These concerns must be taken seriously.

   - Staff responsible for the delivery of social security schemes should ideally be involved during the development of an evaluation design; only by knowing how policies really work or are likely to work in practice, and how they are delivered to potential beneficiaries, can evaluations be made to work effectively.
   - Staff fears can generally be allayed but only if they are convinced that their voice is going to be listened to and, as appropriate, acted upon. This process is often time consuming, requiring tiers of explanation, discussion and negotiation with management, trade unions and individual staff.
   - The aim of evaluators should be to ensure that all those involved in the delivery of social security policy appreciate the importance of the evaluation, have confidence in the process and recognize the significance of procedures and precision. Staff need to know what is expected of them and to know from whom to seek advice and who to inform if things go wrong.
   - There is merit, wherever possible, in consulting and engaging with staff early in the detailed design of an evaluation. Staff are often the best placed to judge on matters of feasibility and through their involvement they may share ownership of the evaluation.
   - Most important of all, staff have to know and believe that it is the policy intervention and not their competence that is being evaluated and that their confidentiality and that of their clients will be respected.

4. *Analysis and dissemination*
   It is vitally important that the results of evaluations are used. Evaluation is time consuming and expensive. Clients will have been put to inconvenience and even at risk of harm for the benefit of collective learning. It is therefore important to maximize the chances of the results being used.

   – Barriers to effective use
     A number of factors often conspire to inhibit the use of evaluation results. These include:
     – Evaluations are often lengthy and policy interest may move on;
     – Governments and key staff may change;
     – The results may be politically "inconvenient";
     – The results may be equivocal or complex;
     – The results may not have clear policy implications or, if they do, political and considerations may constrain their use;
     – The report may be – but should not be – poorly drafted and disconnected from key policy issues.

   – Strategies to promote the use of evaluation results
     A number of strategies can be adopted to increase the likely impact of a policy evaluation. Some apply to individual evaluations. Others relate to building acceptance of a culture of evaluation. The most important strategies to increase the use made of individual evaluations are:
     – Good design to ensure that the findings will be open to clear interpretation and, so far as possible, generalizable;
     – Involving staff engaged in policy delivery in the interpretation of results, inviting them to reflect not only on their understanding of the workings of the policy as a whole, but also of the credibility and salience of the emergent findings;
     – Sharing and taking feedback on early findings and thus preparing the policy community for the results with the prospect of avoiding knee-jerk reactions to unexpected and perhaps unwelcome findings. The more evaluators and policy-makers know about the policy imperatives and the tone of the emergent findings, the more likely it is that the results will be interpreted sensitively and used productively within the policy process.

   – Promoting a culture of evaluation
     There are strong reasons for making the results of evaluations publicly available and open to scrutiny.

     A further level of accountability can be injected into the policy-making process if evaluation is seen to be unbiased and immune from political interference. This, in turn, increases the likelihood that the results will be accepted by all the parties concerned.

     Independence in evaluation is aided, at national level, by adopting similar protocols with respect to quality, peer review and the commitment to publication as apply to national statistics which are often, in turn, subject to international convention.

     In the short term, it is important to ensure that results are presented in a balanced and neutral fashion in the evaluation report, irrespective of the political spin that might accompany publication. There is some evidence that, in the longer term, the development of an evaluative culture is accompanied by the realization that ideological policy-making can constructively be tempered with evidence, that all policies and social security schemes can probably be improved, that few are likely to be complete failures and that the risks associated with testing out social security schemes are generally less than those associated with rolling out or continuing with a flawed policy or scheme.

     Once it is accepted that social security policies and schemes should be evaluated, it is somewhat easier to distance policies from personalities.

## 8.4 Summative evaluation/experimental methods (random assignment)

*Summative*, *impact* or *programme evaluation* (the three terms are interchangeable) makes explicit reference to (a) the baseline situation, and (b) to a counterfactual – the situation that would exist in the absence of the policy or scheme being evaluated.

The goal of summative evaluation is to establish whether a policy or a social security scheme works – that is, whether it meets the performance goals set for it and hence whether it fulfils its objectives. Summative evaluations are frequently accompanied by formative evaluation to establish, amongst other questions, *why* the policy works or fails to work.

While certain designs (e.g. difference in difference and propensity score matching) permit *existing* or *previous* policies to be evaluated summatively, in the fields of social security and social policy summative evaluation is often *prospective*.

*Prospective evaluation* often entails implementing a new policy, typically on a comparatively small (pilot) scale, and simultaneously subjecting the pilot to evaluation.

Summative evaluations are often divided into two types: *experimental* and *quasi-experimental*.

1. Experimental methods allocate individuals at random into either the *programme group* to receive the benefits of the programme being evaluated or into a *control group* which does not. The control group is the mechanism for defining the counterfactual. The difference in the average of outcome between the programme and control groups provides a maximally precise, unbiased estimate of the impact of the policy being evaluated.
2. Quasi-experimental methods use some other means of defining the counterfactual and cannot be relied upon to yield unbiased estimates of programme impact. They therefore tend to be used when randomization is not possible.

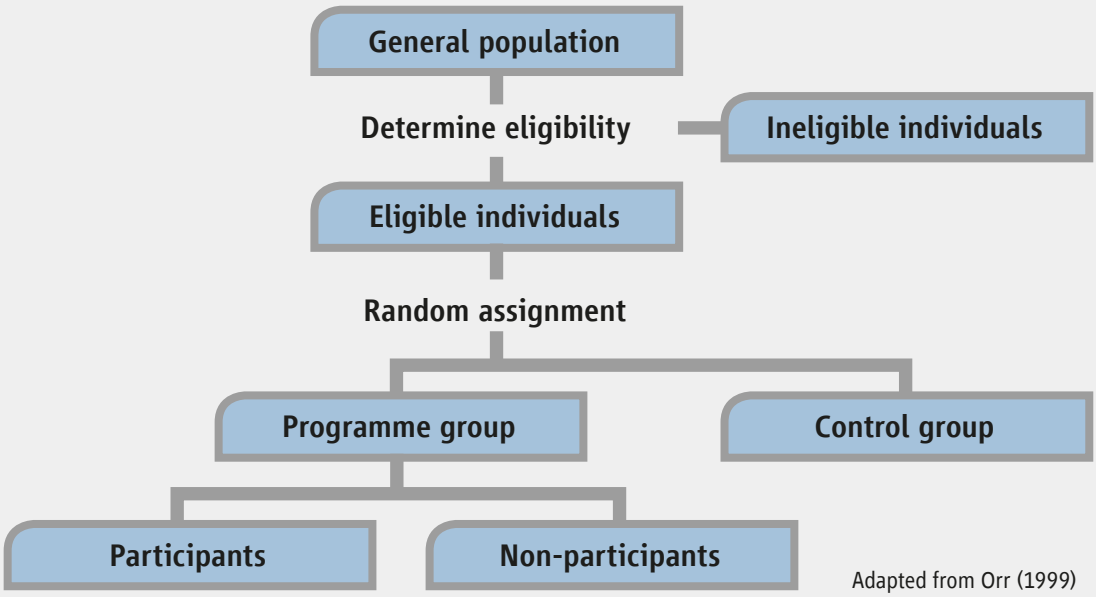
The process of random assignment is illustrated in Figure 8.2. To simplify matters, it is presumed that:

- You want to evaluate a new social security or activation scheme that is targeted on applicants with prescribed characteristics – the eligible population.
- The intention is that the scheme should enhance some aspect of well-being, say, household income.

The following steps are involved:

- Determine who meets the eligibility criteria.
- Randomly allocate those who are eligible to either:
  - the *programme* group in which they receive the benefits and services associated with the new scheme;
  - the *control* group in which they receive no benefits or services.
- Implement the programme for a sufficient length of time for the effects of the programme to materialize.
- Measure and compare the average outcomes (household income) for the programme and control groups:
  - the difference in average outcomes provides a direct unbiased measure of the impact of the programme.

Figure 8.2. Random assignment




Adapted from Orr (1999)

### 8.4.1 Rationale for random assignment

Policy experiments using random assignment are taken to be the "gold standard" against which all other designs are assessed. There are a number of reasons for this:

- Random assignment yields unbiased estimates of programme impacts
  - Randomization ensures that the characteristics of the persons assigned to the programme and control groups do not systematically differ with respect to EITHER *observable* OR *unobservable* attributes.
  - Therefore, any differences in average outcomes observed between programme and control groups can confidently be attributed to the effect of the policy programme. This is because in all other respects the groups are equivalent.
- No other method can guarantee producing unbiased estimates of programme impacts
  - While quasi-experimental designs strive to make sure that programme and control groups are identical in terms of observed characteristics, they cannot ensure equivalence with respect to unobservable ones. Only random assignment guarantees that, sampling error aside, programme and control groups are equivalent and impact estimates are unbiased.
- No baseline measures are required if samples are large
  - If the samples are (infinitely) large, the mean characteristics of randomly assigned programme and control groups are identical. There is therefore no need even to measure outcome variables such as household income *before* the introduction of the policy since these, too, will be the same due to the randomization. This clearly reduces and simplifies the research task.
  - However, if the samples are small (which is most often the case), there may be significant differences in the baseline situations of the programme and control groups which need to be checked for and taken into account in analysis.
- When people fail to participate, it is easy to estimate the impact of the policy on those who do
  - In practice, not all persons assigned to the programme group are likely to receive the benefits or to participate in activities. To the extent that they fail to participate, the apparent impact of the policy will be diluted since those not participating will, in effect, have the same experience as the control group. (See Box 8.2. for a case study of random assignment in Argentina.)
  - It is possible, however, to make a *non-show adjustment* (see Box 8.3.) which permits an estimate of the effect of the programme to be calculated only for those people who actually participated instead of all those assigned to the programme group.

## Box 8.2. The impact of wage subsidies on unemployment in Argentina

**Design**

Random assignment of unemployed persons to two programme groups and one control group.

**Context**

Argentina privatized public sector activities in the late 1990s. This led to a sharp rise in unemployment. "Workfare" was often the main form of welfare assistance. It typically took the form of temporary low-waged work oriented to social infrastructure or community services. There was much concern that this form of provision engendered welfare dependency. Policy attention shifted to consider the possibility of introducing wage subsidies and training to break the pattern of dependency.

**Policy question**

The "Proempleo" experiment was designed to assess the efficacy of a wage subsidy and specialized training in the transition from workfare to regular work.

**Details of design**

Households with workfare participants were identified; 953 answered the baseline questionnaire. They were randomly assigned to three roughly equal sized groups. One received a voucher worth between $100 and $150 per month. One got a voucher and training. One group was reserved as the control group. Workfare participants were free to choose whether or not to participate in the programme offered. All were subsequently re-interviewed at six-monthly intervals.

**Outcomes of interest**

- Private sector employment;
- Self-employment;
- Employment in a temporary employment programme;
- Wage earnings.

**Rationale for design**

The design sought to exploit the power of random assignment while not mandating clients to participate in the programmes being evaluated.

**Findings**

Eighteen months after the baseline survey:

- 14% of voucher recipients had private sector jobs;
- 9% of control group had private jobs;
- This statistically significant difference was largely confined to women and younger workers;
- Very few employers sought to recoup the subsidy which required them to register the employee and hence to incur administrative costs and government social charges.

**Issues**

The experiment raised a number of ethical issues:

(1) It was not announced publicly and informed consent was not secured.
(2) None of the beneficiaries (in any of the three groups) were told that they were part of an experiment.
(3) Ministry staff were not informed of the randomization of beneficiaries.
(4) Despite attempts to ensure that the different options were kept secret at least 40 members of the control group asked to join the programme group but were denied access to it.
(5) Despite randomization, the programme and control groups were not equivalent. A "difference in difference" analysis was used to correct for observed differences between the three groups.
(6) While take-up of the subsidy attained 100%, that for the combination of subsidy and training reached only 70%. Analysts tried to correct for this by a statistical randomizing technique.
(7) The design was not balanced in that it was impossible to estimate the impact of training alone.

*Source: Galasso et al. (2001).*

**Box 8.3. The non-show adjustment**

In random encouragement designs, or in simple random assignment experiments where not everyone actually participates in a programme, the interest is often in establishing the impact of the programme on those who actually received it.

The impact on participants is obtained by dividing the overall impact *(I)* by the participation rate *($r_p$)*:

$$I_p = \frac{I}{r_p} \qquad (1)$$

This can be shown to be the case by recognizing that the overall impact can be obtained from the following equation:

$$I = r_p I + r_n I_n \qquad (2)$$

where $I_p$ and $I_n$ are the impacts on participants and non-participants, and $r_p$ and $r_n$ are the proportions of participants and non-participants in the programme group.

If the impact on non-participants is zero, then the overall impact is given by:

$$I = r_p I_p \qquad (3)$$

Solving equation 3 for $I_p$ yields equation 1.

### 8.4.2 Determining the scale of a random assignment experiment

The cost of random assignment experiments can be curtailed if they are only as large as necessary. In order to determine the appropriate size it is necessary to know a little bit about sampling.

- It is not necessary to involve everyone
  - Although many experiments do seek to reach the entire eligible population, this is not always essential. The results of an experiment can be generalized if a sample of the population is involved provided that the sample is drawn randomly from the eligible population.
- A random sample is not the same as random assignment
  - It is easy to confuse random sampling with random assignment but they are not the same. In a random assignment experiment a *sample* is first drawn *at random* from the eligible population and the sample then *randomly assigned* into programme and control groups.
- Selecting the appropriate sample size
  - The sample size is best determined by policy requirements: What is the smallest policy impact that needs to be detected and with what given level of confidence? This criterion is referred to as the *minimum detectable effect*. Other things being equal, the larger the sample size, the smaller the effect that can be detected. When setting the sample size it is important to consider whether it is important to measure differences between subsets of the population. If so, the size of subgroup together with the size of the effect that you want to measure will set a lower limit to the sample size. (See Orr, 1999, p. 18 for a discussion of the minimum detectable effect and how it can be calculated).

### 8.4.3 Limitations of random assignment

Despite the attractiveness of random assignment, it has at least four strategic limitations and three sets of practical ones.

#### Strategic limitations

It is generally impossible to use random assignment to evaluate:

- An existing policy or scheme: this is because existing provisions would need to be randomly withdrawn from people in order to constitute a control group, which would usually be judged to be unethical;
- Second-order effects of a policy/scheme such as the impact of an activation policy on the wages of people who actually take up employment;

- A policy that is *intended* to affect the system as a whole (as well as individuals within it);
- A policy that may have unintended consequences at a system-wide level (e.g. a work activation policy that could possibly reduce wage levels) (Bottomley and Walker, 1996; Burtless and Orr, 1986).

### Practical limitations

1. *Difficulty of generalizing results*. The ability to generalize from an experiment is called *external validity*. Most random assignment experiments have low external validity. This is because:
   a. It is usually more practical to trial a programme in one or a small number of localities than to experiment on a national scale;
   b. Samples are representative of the localities in which the policy experiments are being run and not of the entire national eligible population;
   c. There is no statistically valid way of grossing up such local impact estimates to the national level.

   The statistical solution is to draw a representative sample of areas and either to randomly allocate areas to implement the policy or to serve as controls or to sample the eligible populations within them and to allocate these to action and control groups. Both strategies often prove to be too difficult or expensive to implement.

2. *Public acceptability*. There is often a reluctance to allocate public services randomly even for the purposes of evaluation. The reasons can be ethical (see below) and emotional. These concerns can become an issue when local social security staff are involved in implementing random assignment:
   a. Staff may not want deny services to clients whom they believe would benefit from receiving them;
   b. Staff may not want to provide services to clients who are not perceived to deserve them;
   c. Staff may not believe that the allocation is actually random;
   d. For any of the above reasons staff may try to circumvent the random allocation.

   It is therefore very important to explain the principles of random assignment to staff and to ensure that they understand its importance. In addition, it is essential to make it impossible for staff to interfere with the random allocation procedure by:
   – Allocating cases by computer;
   – Making random allocation the responsibility of evaluators not local administrative staff.

3. *Implementing random allocation*. There are occasions when it may be impossible for you to use random assignment to evaluate a policy because it is difficult to allocate people to a programme group and a control (see Box 8.4.):
   a. *When there is universal coverage*: It is generally difficult to employ random assignment to a policy with universal coverage since there is nobody available to constitute a control group because everybody must be free to receive the policy. Similarly, it would be difficult to assess the impact of a mass benefit take-up advertising campaign using random assignment because everyone would be exposed to it;
   b. *When the target population is vulnerable*: Ethical objections might be raised if the target for a work activation policy included an unknown proportion of very vulnerable people since these could knowingly be placed at a health or other risk through random assignment. (It nevertheless remains open to question whether this ethical dilemma is truly attributable to the random assignment or to the policy itself; see Walker, 2000);
   c. *When the target group is dispersed*: If the target group is geographically much dispersed, locating and allocating them into programme and control groups would be very expensive.

d. *When a policy already exists*: As already noted it is generally considered to be ethically unacceptable to withdraw existing benefits to create a control group. Usually the existing provision is left in place and made available to the controls. This means that only any *marginal* improvement of the new policy over the old one is evaluated.

### Box 8.4. The benefits of national health insurance in Mexico

**Design**

Random assignment of geographic areas with pair-wise matching.

**Context**

SPS is the new Mexican Universal Health Insurance programme (*Seguro Popular de Salud*). It will ultimately provide medical care, drugs, preventative services and financial health protection to the 50 million Mexicans without health insurance. The introduction of SPS is believed to be one of the largest health reforms of any country in the last two decades. The evaluation was commissioned by the Mexican Ministry of Health (MoH), and implemented by Harvard University.

**Policy question**

SPS is better described as a "social welfare" programme. It provides preventive and regular health care, as well as subsidies to reduce out-of-pocket health expenditures. It targets primarily lower income uninsured Mexicans. It also aims to strengthen the certification and effectiveness of local health facilities.

**Outcome measures**

- Satisfaction with the health-care provider;
- Self-assessments and reports of health, chronic conditions and risk factors;
- Blood pressure, cholesterol, blood sugar and HbA1c;
- All obtained from specially commissioned survey data.

**Details of design**

- There are over 8,000 health clusters in Mexico's 31 states, including an actual or future health clinic or facility;
- 13 of the 31 Mexican states agreed to participate in the evaluation;
- Their 7,078 health clusters (5,439 rural and 1,639 urban) were matched in pairs;
- 74 of these pairs of health clusters were selected from 7 states comprising almost 80,000 households and over 0.5 million individuals;
- One health cluster from each pair was assigned to receive encouragement to individuals to affiliate with SPS, along with the health facilities, drugs and doctors necessary to implement the programme effectively. The other health cluster in each pair received nothing extra;
- At the time of random assignment, baseline surveys were conducted of the health facility within each health cluster, and of about 32,000 randomly selected households within 50 of the 74 pairs of clusters;
- Follow-up surveys of the health facilities and individuals within each health cluster were conducted ten months after random assignment, and then repeatedly at other intervals.

**Rationale for design**

Randomizing of individuals was impossible because:

- it would have been politically and ethically unacceptable;
- every citizen is technically permitted to affiliate even when no health facility is nearby.

However, the SPS was not an entitlement at the cluster level. Some clusters were excluded including those where politicians were insistent that the scheme had to be implemented and areas where high proportions were affiliated with the scheme and where encouragement therefore might not have had much effect.

*Source: King (2007).*

### 8.4.4 Added sophistication

There are many ways in which the basic random allocation design can be adapted to address some of the limitations noted above and to cope with more complex scenarios.

- Adjustments for small sample sizes
  - It is rarely possible to have samples large enough to avoid the need to take baseline measures or to control statistically for differences in the characteristics of the programme and control groups.
  - The usual method of controlling for these differences is to use *multiple regression analysis* (or analysis of covariance) (see Box 8.5 below). This permits you to establish the difference in average outcomes between the groups attributable to the policy under evaluation having controlled for variations in observable group characteristics.
  - Such an analysis requires the collection of additional information from or about group members. You should collect information that you believe is most likely to mediate the impact of the policy being evaluated. Often the most important variable in this regard is people's *baseline score* on the outcome variable (for example, household income prior to the introduction of a new benefit).
- Coping with an existing policy
  Very often a policy is already in place and you need to evaluate the marginal impact of the new policy. This adds complexity because:
  - Recruitment into the experiment has to be integrated into the existing policy intake process. This must be done to minimise the effect on the composition of the participant population as might happen if, for example, an additional interview or step in the intake process were to be imposed.
  - It may be necessary for you to solicit the informed consent of potential participants to participate in the policy experiment prior to random assignment. This is an ethical consideration sometimes reinforced by legal obligation. However, for instance in the USA, the need for consent has been waived for mandatory government programmes. There it is presumed that eligibility for receipt of a programme simultaneously entails the obligation to participate in any approved evaluation.

Figure 8.3 illustrates the need to introduce an invitation to participate after initially assessing eligibility for the programme being evaluated.

## Box 8.5. Regression adjusted impact estimates

Where there is no equivalence between the programme group and the control it is advisable to try and control for all the differences that can be measured. The commonest approach used to do this is through *regression analysis*.

In regression analysis, a variable of interest, the so-called dependent variable, is described as the linear combination of a set other variables (called independent variables).

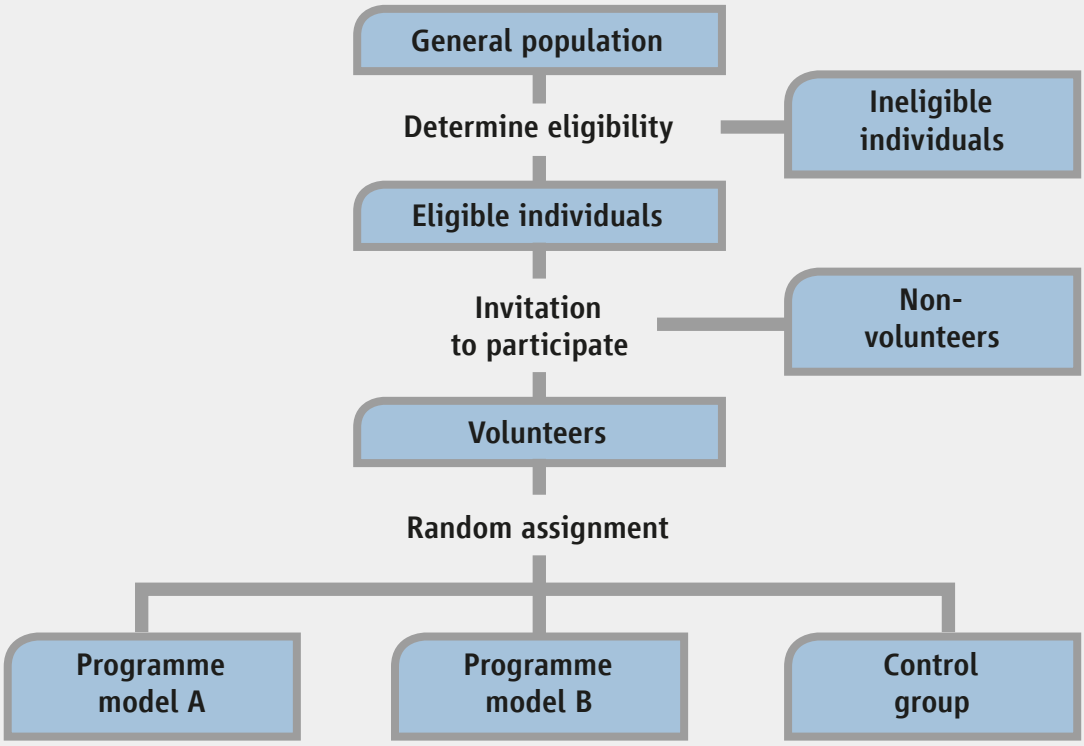
Imagine that all people in the programme and control groups are taken together. In the following equation,

- $Y_i$ is the outcome measure for the ith individual
- $X_{ji}$ is a background characteristic ($j$) of individual $i$
- $P_i$ indicates whether individual i is a member of the programme group (when $P_i$ takes the value 1) or a member of the control (when $P_i$ takes the value 0).
- $e_i$ is a random error term and $a$ is a constant: $Y_i = a + cP_i + b_jX_{ji} + b_kX_{ki} + b_nX_{ni} + e_i$

In this model the coefficient c – usually called the regression-adjusted impact estimate – measures the impact of the programme having controlled for all the measured differences between the characteristics of people in the programme group and those in the control.

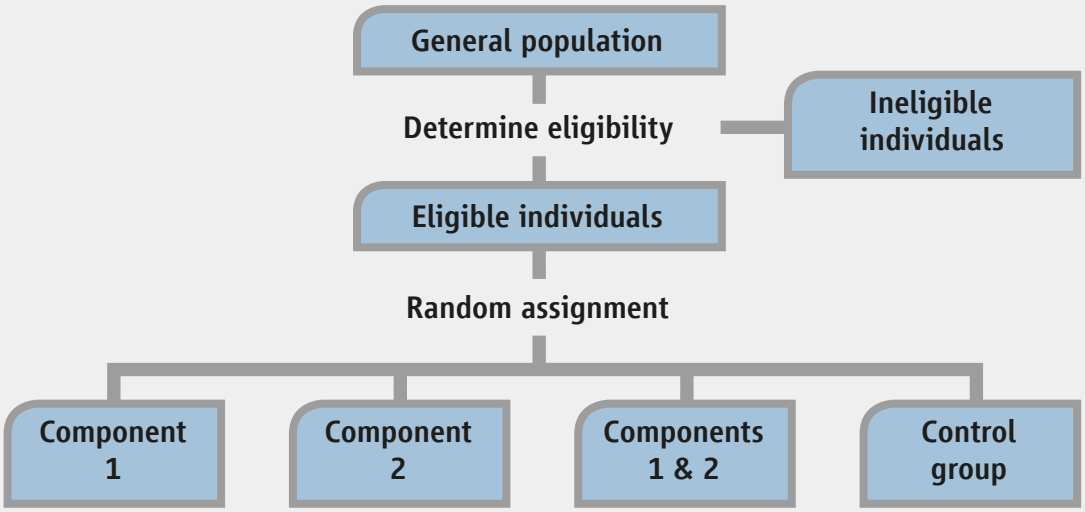
The other coefficients ($b$) measure the effect of the background variables on the outcome measure.

## Figure 8.3. Voluntary random assignment to competing policy model




Adapted from Orr (1999)

Figure 8.4. Random assignment with different policy components




Adapted from Orr (1999)

### Competing policy designs

You can use experimental methods to test the relative effectiveness of competing policy designs.

Box 8.6. illustrates an experimental design to compare the effectiveness of two policy packages directed at the same target population. Note the following:

- It includes an eligibility assessment and an invitation to participate; the latter is likely to be required in most jurisdictions for ethical reasons (see below);
- All volunteers are randomly assigned either to one of two policy interventions or to a control group (either no intervention or the existing policy);
- The effectiveness of each option is determined by comparing the outcome for the appropriate intervention group with that for the control;
- Omitting the control group would allow comparisons of the relative effectiveness of the two interventions but not whether either was superior to the status quo;
- If there is no control group *and* some volunteers do not participate it is not even possible to assess the relative effectiveness of the two interventions; a control group is necessary to make the no-show adjustment (see Box 8.3. above).

Figure 8.4 presents an experimental design to assess the specific contribution of particular elements in a policy package, to its overall impact. You might, for example, be interested in the relative contribution made by training and employment subsidies in an activation programme. Note the following points in this so-called *factorial* design:

- Random assignment is into one of three intervention types (e.g. training, subsidy and a combination of training and subsidy) and a control group;
- The effectiveness of each policy component singly and in combination is determined by comparing the outcomes for the respective intervention group with that for the controls;
- It is generally not the case that the effectiveness of an intervention is simply the sum of its parts; if this were the case the outcome for the training and subsidy groups could simply be added together to give their combined impact;
- In practice, policy components usually *interact*, enhancing or reducing the effectiveness of the components;
- The degree of the interaction between components can only be assessed with a full factorial design that includes all possible policy combinations.

### Box 8.6. The value of providing business education with micro-credit in Peru

**Design**
Random assignment of banks with competing policy models.

**Context**
This study was implemented with FINCA Peru, a long-standing microfinance organization. FINCA's mission is to improve the socio-economic situation of poor people and to empower women through the promotion of the village-banking methodology. Typically, FINCA provides loans in cycles of 4 months to groups of 30 women to help them expand their individual small businesses.

**Policy question**
The effect of providing training alongside micro-credit.

**Outcomes of interest**
- Loan repayment: If businesses generate increased revenues, repayment may improve;
- Loan sizes and savings volumes: If clients manage cash flows better, they may need less credit, and loan sizes could decrease;
- Household decision-making: Improved businesses could empower female micro-entrepreneurs with respect to their husbands/partners in business;
- Child labour: Changes to the household enterprise could cause children to contribute their labour to the family business or result in a shift in family priorities to the education of children;
- Client retention: The training could be perceived as an additional benefit of membership in the community bank, and client retention could improve.

**Policy models and design**
Two were compared against a control:
- Mandatory treatment (Lima: 49 banks; Ayacucho: 55 banks). All clients were required to arrive early or stay after the meeting to receive training. Clients were fined for missing training. Continued absence could lead to expulsion from the bank;
- Optional treatment (Ayacucho: 34 banks). Attending training was voluntary;
- No training (Control) (Lima: 50 banks; Ayacucho: 51 banks). The control group clients did not receive training, but continued to have their regular credit meetings as they had in the past.

**Findings**
- Loan repayment among treatment groups was 3% higher than among control groups;
- The treatment groups demonstrated greater business knowledge and better business practices;
- They had sales in the month prior to the survey of 16% higher than control groups;
- However, there was no impact on household decision-making processes although female children dedicated more time to schoolwork.

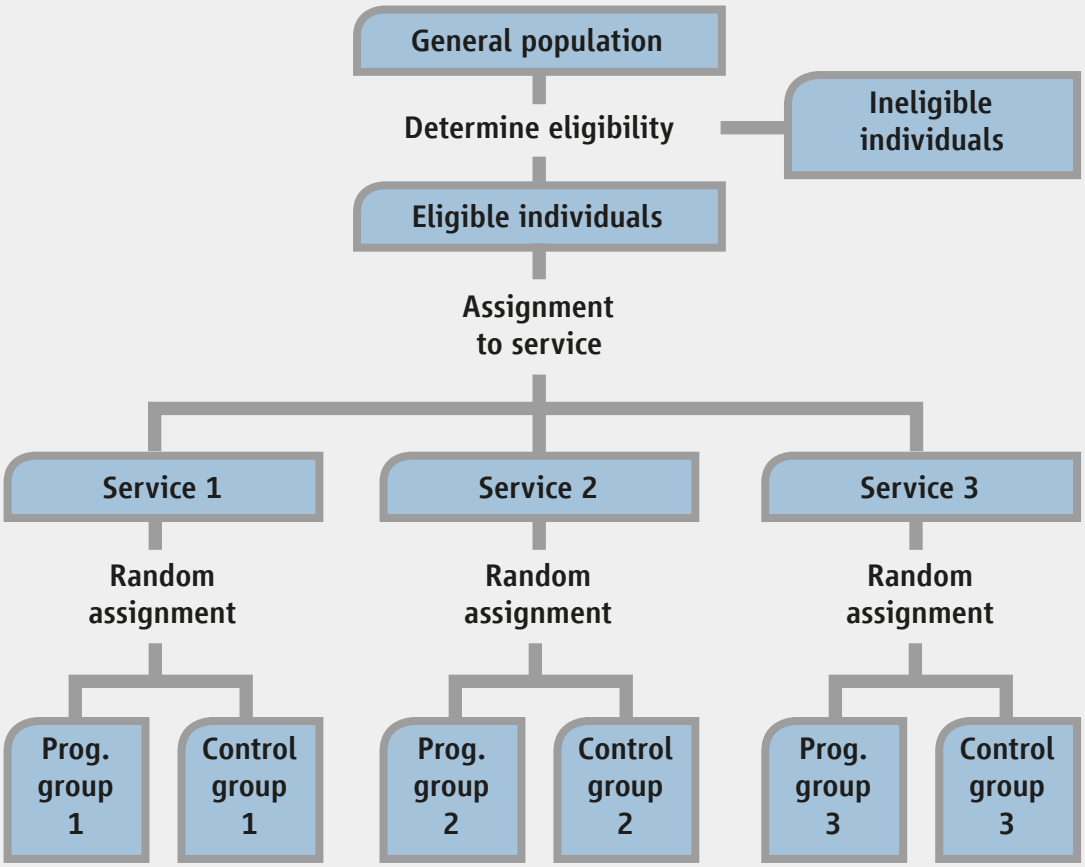
**Sources:** *Karlan and Valdivia (2006); Karlan et al. (2006).*

### Targeting different groups

Very often social security and activation policies are targeted on different groups and adapted to meet their varying perceived needs. It is possible for you to test the effectiveness of such a strategy by adopting a design similar to that presented in Figure 8.5. Note that:
- Three separate policy packages are being evaluated, each tailored to different kinds of participants;
- Agency staff first allocate applicants to the programme that they judge (perhaps in negotiation with individual applicants) to be the most appropriate;

- Applicants assigned to each programme are then randomly assigned to the programme or to a control group;
- Comparison between the programme and control groups provides a measure of the impact of each intervention for the appropriate target group;
- Comparison between the average outcome for the three programme groups and the three control groups indicates the overall effectiveness of the programme;
- However, comparison between the different programme groups has little value since the composition of the three groups is likely to differ systematically according to the judgement of staff assigning applicants to the most appropriate policy package.

Figure 8.5. Random assignment with discrete service components




Adapted from Orr (1999)

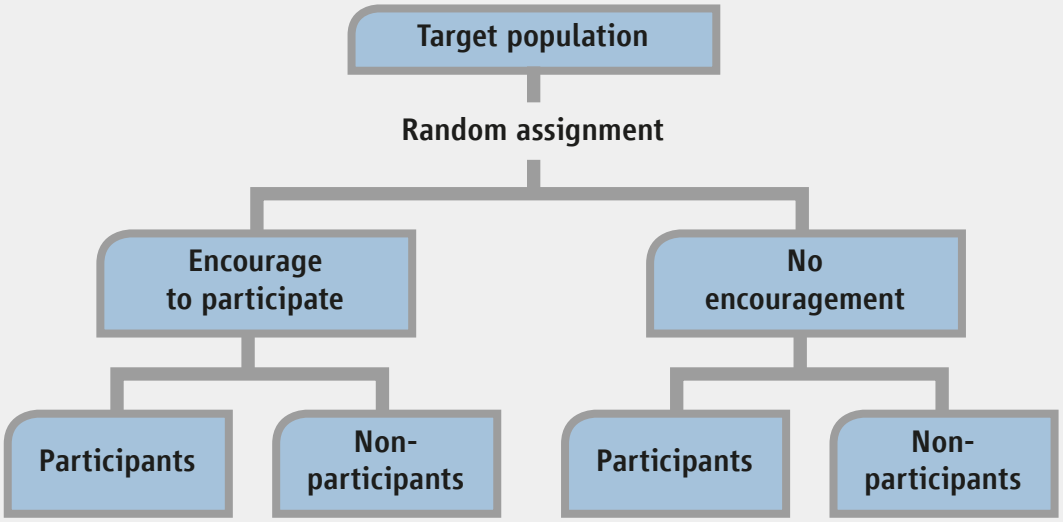
### Avoiding denying benefits

There are occasions when, for ethical, political or practical reasons, it might be considered inappropriate to deny people access to a programme in the way that assigning them to a control group does. This need not mean that random assignment is impossible. Figure 8.6 illustrates a randomized encouragement design and the following points should be noted:

- The target population is defined and randomly allocated into one of two groups in which people are either:
  - encouraged to participate, perhaps by a letter or telephone call; or
  - not encouraged to participate;
- The programme is open to all comers, those encouraged to participate and those not so encouraged;
- The difference in the proportions of people encouraged and not encouraged who take up the programme is established;

- The difference in outcomes is established;
- The impact of the policy can then be calculated as the ratio of the two differences;
- For the randomized encouragement design to work well certain assumptions need to be fulfilled:
  - the encouragement must affect participation in the programme;
  - the encouragement must not directly influence the outcome;
  - those responding to the encouragement would not otherwise have participated.

Figure 8.6. Randomized encouragement design




Adapted from Orr (1999)

## 8.5 Summative evaluation/quasi-experimental methods (non-random assignment)

It is not always possible or appropriate to use random assignment. In such circumstances it is still possible to select from a range of alternative methods that employ a counterfactual.

These "quasi-experimental" methods are generally distinguished by, and named after, the means of defining the counterfactual. Because none of these methods is able to take into account unmeasured differences between programme and control groups, the estimates of impact are susceptible to bias.

A selection of quasi-experimental designs is presented below.

### 8.5.1 Non-equivalent groups and difference-in-difference designs

The hundreds of quasi-experimental designs that fall under this characterization are distinguished by the nature and method of comparison.

#### Nature of comparison

When random assignment is inappropriate, the obvious alternative is to compare two ostensibly similar groups of people, one that receives an intervention, the other that does not. Without random assignment, the groups cannot be equivalent, hence the name of the approach. However, the more similar the groups are, the more likely it is that any differences in outcome will be due to the impact of the policy.

Comparison groups may be defined in many ways:

1. *Comparing areas*
   Comparing geographic areas is a common strategy when a new policy is to be evaluated (i.e. prospective evaluation). There are number of advantages to the approach and a major disadvantage:
   **Advantages**
   – The whole population in the area can be targeted, giving a good approximation to full policy implementation;
   – Bias due to queuing and partial equilibrium can be avoided;
   – No one locally need be excluded from the policy;
   – No one in the control area need know that they are part of a control (though see the section on ethics below).
   **Disadvantage**
   – Geographic areas vary enormously, in themselves, and in their influence on the lives of inhabitants. This makes it very difficult to find areas similar enough to be confident that differences in outcome are truly attributable to the impact of the policy.

2. *Comparing administrative offices*
   Sometimes social security offices are synonymous with a geographic area and the same advantages and disadvantages apply. The principal additional advantage is that of administrative convenience.

3. *Comparing by employment status*
   Employment status can sometimes be used in retrospective summative evaluation of existing social security schemes. Imagine, for example, you wanted to evaluate an activation policy targeted on unemployed men. Possible comparison groups could include:
   – Unemployed women;
   – Men receiving incapacity benefit, the rationale being that in many countries there is a considerable overlap in the skills and employment histories of people claiming unemployment and incapacity benefits;
   – Men employed part-time.

   The disadvantages of these comparison groups are self-evident; each differs from unemployed men such that they might behave similarly.

4. *Comparing by demographic status*
   Like employment status, demographic status can sometimes be used to create a counterfactual by which to evaluate an existing policy. For example, a policy designed to increase the coverage of pensions among prime aged men might take any of the following as possible comparison groups but each has its own limitations:
   – Prime aged women;
   – Men just under prime age;
   – Men just over prime age.

### Method of comparison

The non-equivalence of the groups makes it unlikely that a simple comparison of outcome variables would generate a good estimate of the impact of the policy. This is because:

- The groups will probably differ in value of the outcome variables measured before the intervention which is likely to affect scores afterwards;
- The groups may differ in characteristics that could affect the impact of the evaluation.

## Box 8.7. The effects of conditional cash transfers in Brazil

**Design**
Matched areas design

**Context**
Although it is illegal for children under the age of 14 to work in Brazil, the law has not been effectively enforced. A child labour eradication programme, Programa Erradicacao do Trabalho Infantil (PETI) provided income transfers to poor households in exchange for an agreement that a child would attend school at least 80% of the time. PETI was implemented in poor rural states (Pernambuco, Bahia and Sergipe) in Northeast Brazil. The programme required children to attend an after-school programme that effectively doubled the length of the school day.

**Policy question**
Policy interest extended beyond the question of whether PETI would curtail child labour and enhance well-being to include possible negative effects on children not engaged in the programme who might be expected (tempted) to substitute the hours of work forgone by participating children.

**Details of design**
Six municipalities were selected in each state, three to operate the program me and three to be controls. 200 households with at least one child aged 7–14 were drawn randomly in each municipality and information collected on household, parental and child characteristics. The design produced four groups:
- A1: children who participated in PETI;
- A2: children in the same households as A1 but who did not participate;
- B: children in PETI areas (but not in A1 households) that did not participate; and
- C: children in the control areas.

**Outcome variables**
- school enrolment;
- labour participation;
- hours of work;
- sector of employment;
- highest school grade attained.

**Rationale for design**
Evaluation was planned after the programme started and it was not possible to randomly allocate the municipalities into treatment and control groups (Rawlings and Rubio, 2003).

**Findings**
1. Children who participated in PETI
   - spent more time in school;
   - less time at work;
   - less time in risky work;
   - progressed in school at a faster rate.
2. However, in two sates the PETI programme appears to have increased work hours for non-participant children to work more than 10 hours per week and to have had slightly greater difficulty progressing to the next grade.
3. These adverse effects were not large enough to outweigh the positive effects on participating children.

**Issues**
Because outcomes were measured at one point in time only (ruling out the possibility of a difference-of-difference analysis) the design was very sensitive to the municipalities excluded from the programme serving as adequate controls (after statistically accounting for observed differences).

*Sources: Rawlings and Rubio (2003); Yap et al. (2002).*

The strategy generally adopted to disentangle the effect of the intervention is called *difference-in-difference* or *double difference*. It is illustrated in Box 8.8.:

- Outcome variables are measured before (at time $T_1$ – the baseline) and after (time $T_2$) the intervention for the control group and the differences calculated;
- The same procedure is followed for the programme group;
- The two sets of differences are compared and the difference between them provides a measure of the impact of the intervention;
- While this comparison corrects for differences in the original outcome variables it does not take account of differences in other group characteristics;
- Therefore, as in the case of random assignment experiments with small samples, multiple regression analysis (or analysis of covariance) is used to account for differences in the characteristics of the programme and control groups;
- However, unlike random assignment this procedure results in biased estimates of the impact. This is due to the combination of measurement error of the outcome variables when measured before the intervention;
- The analysis therefore needs to include a correction for the lack of reliability and, because there are different measures of unreliability, slightly different estimates of programme impact are likely.

### Box 8.8. The difference-in-difference design

(also known as: *"non-equivalent comparison group design", NECG)*

| | | |
|---|---|---|
| Description | • A control group can be selected from individuals (groups, areas, etc.) who share some characteristics with the intervention group (e.g. eligibility for the programme, income level, or annual budget).<br>• The two groups are measured before and after the programme.<br>• The control group that did not participate in the programme is measured for 'natural change'.<br>• The intervention group is measured for "natural change" plus change due to the programme.<br>• Subtracting the difference for the control group from the difference for the intervention group gives an estimate of the change due to the introduction of the programme or policy. | |
| Pros & cons | • Can be an easier selection process than other designs with more rigorous selection standards. | • Difficult to attribute results to the intervention and not confounding or spurious variables.<br>• Lack of longitudinal information. |
| Variations & techniques | (a) Establish several standardized post-test follow-up points so as to collect longitudinal data on outcomes. | |

*Source: Adapted from Morton (2008).*

### 8.5.2 Propensity score matching

Propensity score matching is a sophisticated variant of a non-equivalence design that tries to create a control group that is very similar indeed to the programme group.

- The logic of matching
  Matching involves finding a set of people with the same characteristics as those receiving the benefit or intervention that is being evaluated who do not receive it themselves. This can be done by:
  - Matching on an individual by individual basis according to key socio-demographic variables; this is comparatively simple but misses many variables that could influence the effectiveness of a policy;
  - Matching on an individual by individual basis using as many socio-demographic variables as possible; however, the more the variables used, the more difficult it is to find individuals who are a perfect match;
  - Matching by group characteristics, for example by trying to ensure that group averages are the same on key variables; this eases the problem of finding good matches but fails to take account of the interaction between variables.

  Matching works best when:
  - There is a very considerable overlap in the characteristics of people receiving the benefit or intervention and those who are not;
  - Access to the intervention is more or less random for the persons who are potentially eligible;
  - The are a large number of variables available to use for matching;
  - The matching variables can be precisely measured and there are very few missing values;
  - The matching variables are stable;
  - There are large numbers of potential matches to serve as controls.

- Propensity scores
  Propensity scores are a form of individual level matching that helps to address the difficulty of finding exact matches by assigning individuals a single score based on a combination of variables (see Box 8.9). It involves the following steps:
  - Identify the relevant population and obtain/collect individual level data;
  - Identify who has received (or is receiving) the benefit or intervention in retrospective evaluations;
  - For prospective evaluations, if possible, identify who will receive the evaluation;
  - Devise a statistical model that best predicts the people who receive (or will receive) the intervention on the basis of the individual level data collected; this is usually achieved using logistic regression;
  - Use the model to generate a *propensity score* – that is, a measure of the likelihood of receiving the intervention – for everybody irrespective of whether they (will) receive the benefit or not;
  - Take the propensity score for each recipient of the benefit or intervention and find the non-recipient with the most similar propensity score to become their individual control;
  - Compare the mean outcomes for the recipients and matched non-recipients; this is usually undertaken by adopting a difference-in-difference approach.

- Strengths and limitations
  Propensity score matching is an increasingly popular technique. This is because it is:
  - Flexible in that it can be used in many settings and for prospective and retrospective evaluations;
  - Intuitive;
  - Simple to implement;
  - Maximizes the use of available data.

## Box 8.9. What best gets the unemployed into employment in Romania?

**Design**
Retrospective evaluation using propensity score matching of individuals.

**Context**
Employment services (ES) and small business assistance programmes (SBA) were the first major active labour market programmes (ALMP) to be established in Romania after the 1989 revolution. Both were introduced in 1997 and are delivered by public or private service providers under contract to the National Agency for Employment and Vocational Training.

**Policy question**
The aim of the evaluation was to establish the relative effectiveness of the two ALMP programmes.

**Details of design**
A random sample of 3,400 persons registering at the Employment Bureau during 1999 was interviewed during January and February 2002. About two-fifths of this sample were ALMP participants. The others – the potential comparison group – registered at the Employment Bureau but did not participate in an ALMP.

Propensity scores to select a group of participants for ***each*** treatment group were derived using a probit model for the choice between the two programmes and non-participation.

*Outcome variables*
- Re-employment probabilities;
- Workers' earnings in their new job:
  – at the time of the survey;
  – during the two-year period prior to the survey.
- All persons in the sample were unemployed at the baseline in 1999.

**Rationale for design**
Both programmes were already fully implemented but the fact that participation in ALMP was not universal provided the basis for constructing a comparison group.

**Findings**
- ES and SBA had positive impacts compared with non-participation in ALMPs;
- ES was more effective than SBA in helping unemployed people back to work quickly and would even have been beneficial for SBA participants;
- ES improved job-matching but would have been of less value to people with good informal contacts;
- SBA generated human capital benefits for rural workers and those with low-educational qualifications.

**Issues**
- Although the matching appears effective, the programme groups and control are not necessarily equivalent on unobserved characteristics;
- One particular issue here is the discriminating nature of the judgements made by officials in terms of allocating clients to the ES and SBA programmes;
- The cost-benefit analysis is weakened by a failure to take account of potentially significant benefits (e.g. intensified job search prior to entering the programmes in order to avoid participation) and additional costs (e.g. the deadweight loss of taxation to finance benefits, subsidies, and operation of programmes).

*Source: Rodriguez-Planas (2007).*

Nevertheless, propensity score matching still has considerable limitations:
- It is premised on the assumption that cases who did not receive the benefit or intervention did not receive it at random;
- Like other techniques, it can only match on observable and available variables allowing for bias to be introduced by unmeasured characteristics;
- It works best when matches are drawn from the same neighbourhood or locality because otherwise unobserved contextual variables can introduce considerable bias;
- The matching technique tends to exclude individuals with unusual characteristics but which are different in the programme and control groups;
- The size and direction of bias cannot be readily assessed;
- The technique requires quite large numbers of potential controls.

### 8.5.3 Regression discontinuity designs

Regression discontinuity designs are a variant of quasi-experiments using non-equivalent groups that have a distinct advantage: they can be used when you want to target a benefit or intervention on people who need or deserve it most.

- Rationale and approach

  With regression discontinuity designs, persons are allocated to the programme and control groups on the basis of a measure of *need* or *eligibility*. The assignment depends on whether an individual has a score on the measure that is above or below a predetermined, and rigorously applied, threshold or cut-off point.

  Unlike other methods therefore, there is no presumption that the programme and control groups are equivalent prior to implementation of the intervention. Instead, the principal assumption is that later scores would be equivalent were it not for the effect of the intervention.

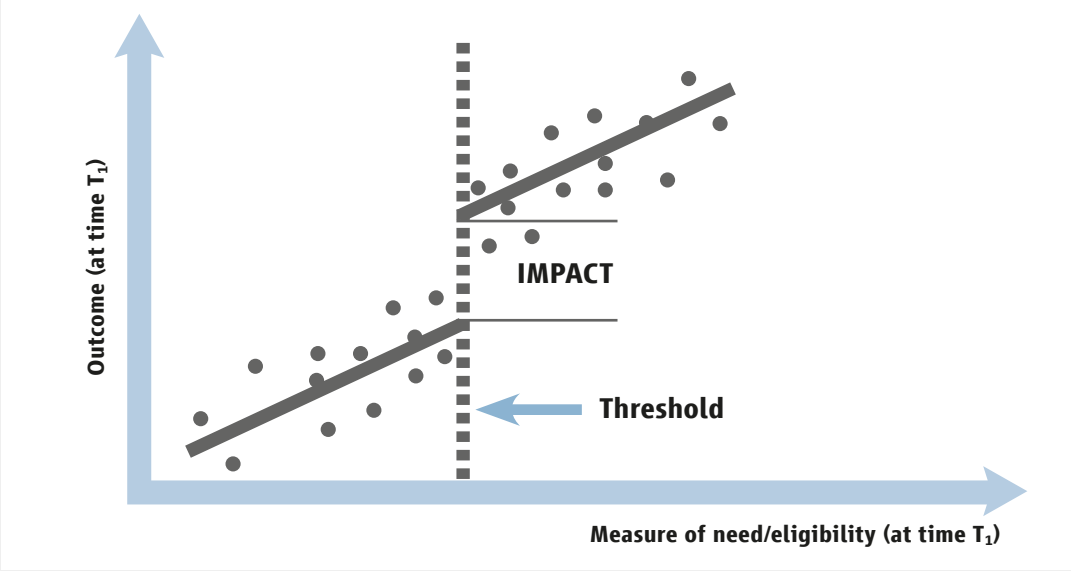
  To implement a regression discontinuity design, you would need to follow the following steps:
  - Determine the target population;
  - Decide on a measure of need or eligibility;
  - Collect or acquire data necessary to make an assessment of need or eligibility for each individual in the population (or a random sample from the population);
  - Fix a threshold or cut-off point; this might be determined on any of a number of criteria:
    - Substantive need, perhaps choosing the poverty threshold;
    - Financial resources, determined by the level of resources available and the proportion of the population you could afford to cover;
    - Statistical robustness, fixing the threshold to ensure sufficient numbers to attain the minimum detectable effect.
  - Deliver the benefit or intervention to those above or below the threshold as appropriate;
  - Measure and compare the outcomes for recipients and non-recipients.

- Analysis

  The analytic strategy is slightly more complicated than with other designs. This is because the assumptions are stronger and the risk of bias somewhat greater. Figure 8.7 illustrates the analytic approach graphically:
  - There is a relationship between the pre- and post-intervention scores that shows a marked break at the threshold. This break is a *discontinuity* in the *regression* line and is prima facie evidence that the intervention had an effect on outcomes;
  - However, such a discontinuity could be caused by other factors, too;

- An early important step is to visually inspect the graph to see whether a discontinuity is apparent. If not, the impact of the policy is unlikely to be great;
- Next you should consider whether the true relationship between the pre- and post-intervention scores might be curvilinear (and, if so, how many "kinks" there might be in the line) and whether there might be any interactions;
- The next step is to develop a statistical model with both higher-order terms (to model a curvilinear relationship) and interaction terms and to test for the simplest model that fits the data adequately;
- If the discontinuity is real, and the model and the data adequate, you should observe a statistically significant coefficient associated with whether or not a person was subject to the intervention. This will indicate the impact of the intervention on the outcome variable (provided the data have previously been adjusted to make the intercept value equal the threshold score).

Figure 8.7. Regression discontinuity design




- Strengths and limitations

  As already indicated, the great strength of the regression discontinuity design is that the intervention can be *targeted* on the most needy while only the less needy are excluded and their exclusion can be justified on principled grounds.

  There are, though, assumptions and limitations that need to be taken into account:
  - The measure of need or eligibility and the threshold or cut-off point must be *rigorously* applied;
  - Ideally the intervention should be uniform in its delivery;
  - The measure of need should preferably be continuous and there must be substantial variation in the level of need among the non-recipient group, in particular, to facilitate robust modelling;
  - The relationship between the measure of need and the outcome measure has to be describable as (or transformed into) a polynomial function otherwise the model will be mis-specified and estimates of the programme impact will be biased;
  - The analysis results can be sensitive to cases around the threshold or cut-off point since they constitute the extreme values among recipients and non-recipients. Depending on the specification of the measure of need, they will constitute extreme high values for recipients and low values for non-recipients, or vice-versa.

### 8.5.4 Summative evaluation without counterfactual

Experimental and quasi-experimental methods all make direct use of a counterfactual. There are occasions when it may be justifiable not to make explicit use of a counterfactual. Estimating the contribution of social security benefits to household income or to lowering the poverty rates are cases in point. Typically, summative evaluations without counterfactuals exploit existing data and are often used to evaluate existing rather then new policies (see Box 8.10.).

Several factors can determine the impact of social security on household income:

- The amount of the benefit;
- Whether the benefit is means-tested, in which case market, or pre-benefit, income will influence the amount of benefit paid;
- Take-up, that is, whether or not an eligible person claims the benefit or, for a range of other reasons, actually receives it;
- Whether receipt of benefit affects other sources of income, for example, either by reducing the work effort (as a substitution effect) or by making it possible to earn more by, for example, subsidizing child care.

The first three factors can be meaningfully investigated without recourse to an explicit counterfactual. The fourth presents much more difficulty.

- First-order effects
  Presented with the requirement to find out how much benefits contribute to households' income, you could undertake a survey of household income. This could provide information on:
  - Benefit income;
  - Market and other non-benefit income;
  - Household and individual characteristics relevant for calculating entitlement to benefit.

  With this information you could calculate:
  - Total income for each household by adding benefit, market and non-benefit income;
  - The proportion (or "increase") in income attributable to benefit;
  - The proportion of people lifted out of poverty through receipt of benefit;
  - People's eligibility and entitlement to benefit and the take-up rate;
  - The benefit income forgone through low take-up;
  - The increase in household income that would occur if take-up increased.

  Although no reference is made in this process to a counterfactual, steps one, two and six all involve an implicit comparison with the counterfactual of a world in which the benefit in question does not exist. In the analysis, each person serves as their own counterfactual: what would their income be if they did not receive the benefit?

- Second-order effects
  While this analysis is valuable, and undertaken as a regular annual exercise in many countries, it does not progress knowledge very far. What if you wanted to know whether benefits increased household incomes after allowing for the possibility that benefits might create work incentives? What if you wanted to know what effect benefits had on other behaviours, e.g. food or alcohol consumption?

## Box 8.10. Large cash transfers to the elderly in South Africa

**Design**
Evaluation without a counterfactual.

**Context**
South Africa supports a comparatively generous old-age pension. It stems from attempts under the apartheid regime to protect white workers without access to occupational pensions. The extension of the social old-age pension to the entire population occurred gradually and reached fruition only in 1993. In the 1990s, when the evaluation was undertaken, the benefit was large equating to twice the median per capita household income of Africans and 3.5 times the $1 per day poverty line.

**Policy question**
In the context of concerns about the distortionary effects of pensions, the evaluation sought to investigate the distributional consequences of the pension (a primary objective of the scheme). It also explored some of their possible behavioural effects of the pension, namely its impact on spending on food, schooling, transfers and savings.

**Details of design**
The project was a secondary analysis of a national survey of 9,000 households conducted in 1993 by the World Bank and the Cape Town University. Each household served effectively as its own control. Income prior to the receipt of social pension was defined as total income less pension income. Sophisticated statistical techniques were used to overcome the resultant biases.

**Rationale for design**
Social pensions were already universally available in 1993 making the specification of a formal counterfactual very difficult.

**Findings**
The research concluded that social pensions were very effective in targeting low-income households. The reason is that take-up was highest among the poorest people while the means-test served mainly to exclude white elders. Pension income was spent in much the same way as other income but it gave greater autonomy to the elderly recipient who, without a social pension, might have had less influence on household decision-making.

**Issues**
The distributional consequences of social pensions were derived purely from arithmetic calculations. It is not self-evident that, in the absence of social pensions, incomes net of social pensions would have been the same. For example, it is possible that without pensions, elderly people would have been forced to seek more paid employment. In other words the counterfactual is poorly specified (an individual being his or her own counterfactual.

This problem is evident, too in comparisons of spending – the disbursement of pension income. Pension recipients were deliberately expected to be different from non-recipients – the rationale for targeting – whereas in the analysis they are the same people.

A further issue linked to the lack of a secure counterfactual is that if pensions are a function of household characteristics and circumstances and those characteristics influence behaviour, then the behavioural consequences of benefit receipt are confounded.

*Source: Case and Deaton (1998).*

These questions draw attention to the weakness of the counterfactual employed in the above analysis:

- The point in time comparison, with each person providing their own counterfactual, is very weak. It assumes that a person's non-benefit income would be the same irrespective of the existence of benefit. Yet, in practice, the non-benefit income is measured once only at a time when benefits already existed. It is at least possible that some benefit claimants previously cut their work hours, using benefits to make up the shortfall in income;
- Benefits tend to be targeted on particular kinds of people, for example, the poor. This suggests that recipients and non-recipients are likely to be different in other ways, educational level, health status, etc.;
- To the extent that the characteristics that make people eligible for benefit are associated with specific patterns of behaviour, it becomes impossible to disentangle whether an observed behaviour, consumption of alcohol for example, is a consequence of benefit receipt or one of the reasons why they needed to claim benefits.

These are real difficulties but they can be tackled in a number of ways:

- Continue without an explicit counterfactual and introduce an increasing number of controls that, in turn, are based on assumptions that may or may not hold good;
- Resort to dynamic micro-simulation models that use individual level data as a basis for simulating what might be expected to happen in given scenarios on the basis of theory and more or less realistic assumptions about human behaviour (that are not always made explicit);
- Revert to an experimental or quasi-experimental format which, if practical, generally provides a much better prospect of generating robust second-order impacts. Practicality – and cost – are deterrents and it may not always be possible easily to evaluate extant policies and programmes.

## 8.6 Cost-benefit analysis

Cost-benefit analysis comprises a set of techniques widely used in appraising policies, often ahead of implementation but sometimes *post facto*. As the name suggests, the approach entails summing estimates of all the benefits (and dis-benefits) associated with a policy intervention and comparing the sum with the total cost. If benefits exceed costs, there is a *prima facie* case for proceeding to introduce the project.

The challenge associated with cost-benefit analysis is, of course, to establish all the outcomes and costs and to convert them to a money metric to facilitate comparison. Not all outcomes, such as increased social solidarity or improved individual morale, are easy to translate into money values. Moreover, as the discussion of impact evaluation indicates, attributing outcomes to particular policies is difficult and ideally requires the specification of a counterfactual.

Cost-benefit analysis has a long history and a vast and complex literature associated with it. It suffices, here, to draw attention to some of the links between impact evaluation and cost-benefit analysis (Orr, 1999).

### 8.6.1 Policy impact, benefits and costs

An impact evaluation employing random assignment not only provides one with unbiased estimates of the policy effectiveness of an intervention in meeting its objectives; it also offers:

1. Unbiased estimates of *all* first-order effects that can be measured. All that is needed, therefore, is to anticipate and measure all the effects of a policy to form the bases for assessing the benefits and dis-benefits of a policy intervention.
2. Unbiased estimates of the *costs* associated with the programme; this is obtained by comparing the costs arising from delivering the programmes.

Quasi-experiments similarly offer a basis for estimating the benefits and costs associated with a policy intervention but there will be some bias due to unmeasured differences between action and control groups.

### 8.6.2 Perspectives

The costs and benefits of a policy can be viewed from many different perspectives. Therefore, an early decision should be made as to which one(s) should be considered. The following are often included:

1. *Social security agency*
   Let us assume that the policy is a new social security scheme. Costs incurred by the agency will include:
   – *Administration costs* associated with advertising the scheme, processing claims and assessing entitlement, making payments and promoting financial security;
   – *Overhead costs* associated with management of resources, including personnel and property; and,
   – *Cost of payments* paid out (depending on accounting conventions used).

   It is important to note the distinction between:
   a. gross costs: the direct outlays required to operate the programme;
   b. net costs: the changes in costs as a result of the specific decision. Even with a new programme, net cost may be less than gross costs if, for example, existing staff take on administration of the new scheme as part of their duties without any increase in remuneration.

   Net cost should always be used in a cost-benefit analysis.

   *Start-up costs* include the cost of developing the policy as well as costs incurred in learning how to operate the programme including the costs of refinement. In costing a new programme, start-up costs would need to be included in the cost-benefit analysis. However, in evaluating an existing programme they would generally be ignored and treated as fixed or "sunk" costs.

   The agency may not incur many benefits from implementing the new programme. To the extent that staff are more fully occupied, this would contribute to increasing the administrative efficiency overall. The new policy might reduce demands elsewhere on the administrative system, for example if the new benefit replaced a means-tested policy that was more expensive to administer.

2. *Social security recipients*
   In contrast to the perspective of the social security agency, from the point of view of social security recipients you would expect the benefits of the introduction of a social security benefit to outweigh the costs. The recipients would gain from the receipt of benefit although, as with costs, it is important to distinguish gross and net benefits. For example, the gross payment of benefit might be offset by the reduction in another means-tested benefit.

   Moreover, it is possible that recipients would incur some costs in order to claim the new benefit. Such compliance costs would include *time costs* associated with completing an application and possibly *travel costs* if an interview was required at the social security office. Sometimes social security schemes have other *psycho-social costs* such as the stigma associated with dependency on state payments. Most of the cost-benefit analyses that accompany impact evaluations ignore such non-monetary costs, partly because of the difficulty of assigning them a monetary value, but in doing so they overstate the cost-effectiveness of the programme in social terms.

   It is worth noting that many of the costs to the social security agency appear as benefits to recipients. However, agency costs will generally exceed the payments paid to recipients due to the administration costs associated with delivering the scheme.

3. *Government*
   The costs to government are likely to be different again. It may bear some or all of the costs of the benefit payment and administration.

   However, it is likely to gain more on the benefits side than does the social security agency. To the extent that the social security payment is taxable, it would claw back some of the expenditure through higher tax payments from the welfare recipients. To the extent that the social security payments lift household incomes, they are likely to gain from incurring less costs associated with poverty because of improved health, increased educational attainment, lower levels of crime, etc.

   Taking a political perspective, the government of the day may gain in popularity if the introduction of the social security scheme appeals to the electorate. However, again most cost-benefit analyses would ignore political benefits of this sort and restrict the analysis to consideration of the public expenditure balance sheet.

4. *Society*
   A further perspective is that of society as a whole. The logic underlying the interpretation of cost-benefit analysis is that those policies that generate net benefits to society as a whole are worth pursuing.

   This does not, of course, mean that all sections of society will benefit from such a programme. For example, whereas eligible claimants are likely to gain financially from receipt of social security payments, taxpayers not in receipt of the benefit are likely to be worse off although, conceivably, their financial loss may be offset to some extent by the knowledge that people poorer than themselves have benefited from society's collective generosity. Of course, greater social security for the disadvantaged members of society will often also bring tangible – but difficult to measure – benefits for the more affluent, e.g. reduced social tensions and crime, as well as higher productivity of labour. Often, evaluators know that these kinds of benefits are important, but leave them outside their analysis because of methodological difficulties in documenting such benefits.

### 8.6.3 Accounting for time

The fact that expenditures on a programme occur at a different time from the benefits serves to complicate cost-benefit analysis. Three areas of complexity warrant attention:

1. *Discounting*
   Current expenditure imposes a greater burden than future expenditure because it more directly limits resources available to be spent in other ways. Costs therefore have to be adjusted, that is *discounted*, to take this phenomenon into account. Analogously, future benefits may be less attractive than immediate ones.

2. *Adjusting for inflation*
   Flows of costs and benefits must also be adjusted to take account of *inflation* since the same cash expenditure in two years' time is likely to be worth less than today.

3. *Long-term impacts*
   Most evaluations last a short time but the effects of policy interventions may be enduring. In so far as it is possible, the flows of benefits need to be estimated. Often this is done by exploiting the widespread finding that benefits accruing from policy interventions tend to decay over time. If such a trend can be detected over the life of an evaluation, then the predicted value of the benefits accruing to the policy can be cumulated until the time that they are forecast no longer to accrue.

## 8.7 Formative and retrospective evaluation

Whereas summative evaluation asks whether a policy works, formative evaluation is more concerned with *why it works* or does not work. Formative evaluation focuses more on process than on outcomes and for that reason is sometimes called *process evaluation*.

Retrospective evaluation and monitoring are also covered here because similar methods are often used although quite often the focus of interest is on whether a policy is working or worked in the past.

### 8.7.1 Formative evaluation

For each summative evaluation question, there is a set of corresponding formative ones that seek *explanations for*, or *understanding of*, the outcomes of policy.

These questions lend themselves to qualitative research and the grammatical tense in which each question is asked again influences the evaluative design, though not to the same extent as with summative evaluation:

- *Ongoing* evaluation (asking "How is it working?") tends to rely on some or all of the following:
  - In-depth and group interviews with policy actors including agency staff and benefit recipients;
  - Policy papers;
  - Observational techniques;
  - Action research in which the researchers engage with policy actors in real-time to assist them in refining policy design and implementation;
  - Participative research in which the traditional subjects of evaluative research, for example benefit recipients and policy administrators and staff, individually or together take a varying degree of control of the evaluation. Their involvement can range from consultation by the evaluators about the approach and methodology to taking full responsibility for the direction and execution of the evaluation.
- *Retrospective* evaluation (asking "How did it work?") will use:
  - Documentary evidence;
  - In-depth interviews with former policy actors;
  - Secondary statistical and other administrative sources.
- *Prospective* evaluation (asking "How would it work?") will employ:
  - Projective techniques to pose questions of policy actors of the form "What do you think would happen if …?";
  - Simulation and role-play in which the dynamics of the policy are enacted by various parties to see practically, if somewhat artificially, what might happen;
  - Laboratory experimentation in which individuals are presented with scenarios in carefully controlled circumstances to try to elicit likely behaviour.

The broad approach of most formative evaluation is pluralistic and investigative, assembling whatever relevant information it is possible to obtain, and comparing and contrasting insights gleaned from the perspectives of different policy actors and sources of data.

The *HOW* and *WHY* – questions that are at the centre of formative evaluation always tend to appear as being more subjective or more interpretative than the *WHAT* and *HOW MUCH* – questions of summative evaluation. The illusion of objectivity and rigour that is commonly associated with summative evaluations and quantitative analysis is more difficult to maintain in formative evaluations using qualitative methods. The fact is, however, that quantitative and qualitative analysis are both grounded in numerous, often competing, philosophical traditions. These traditions differ profoundly in terms of:

- *Ontology*: what they believe the world is like; and
- *Epistemology*: how they think they can know about the world and study it.

The differing traditions will often disagree about such profound questions as the following:

- What do we believe about the nature of reality?
- How do we know what we know?
- How should we study the world?
- What is worth knowing?
- What questions should we ask?
- How should we personally engage in the enquiry?

Just two of the traditions, *pragmatism* and *realism*, are discussed below.

1. *Pragmatism*

   Pragmatism seeks to sidestep the issue of ontology and what the world is like by arguing that all that is known about reality is known through *human experience* of it. Pragmatism then offers an epistemological basis for what scientists and evaluators do, namely seek to solve problems identified through their experience. Their success in solving problems justifies their activity and the method, or combination of methods, used.

   Even within the pragmatist framework, there are different approaches and models that reflect the different purposes of evaluation (Patton, 2002):

   – *Objective-orientated model*: A policy is evaluated with respect to the *objectives* set by the policy architects. The evaluation seeks to:
     – Investigate how the policy is implemented;
     – Understand better how interactions between the various policy actors might affect outcomes;
     – Triangulate different perspectives on the working and effectiveness of the policy.
   – *Goal-free evaluation:* In contrast, does not prioritize those outcomes that are directly linked to the policy objectives. Instead, it explores a wide range of *intended* and *unintended* outcomes and the antecedent processes.
   – *Transaction evaluation.* Emphasizes the different perspectives of all the policy actors, policy-makers, administrators, field-level staff, clients etc. It may entail direct interaction between evaluators and some or all of these groups with a view to working collaboratively to improve implementation and outcomes.
   – *Utilization-focused evaluation:* Prioritizes understanding of the policy implementation process in the belief that the way that policy is shaped by delivery critically influences its likely success.

   Pragmatic formative evaluation has a number of applications:

   – *Supporting impact evaluation*. This is perhaps the most common application when formative evaluation accompanies an impact evaluation based on an experimental or quasi-experimental model.

     In this scenario, the qualitative approach is used to help interrogate, interpret, illuminate and illustrate findings from the impact evaluation. Among the specific questions that might be addressed are:
     – How is the policy delivered in practice?
     – How do clients enter the system and what happens to them as they progress through the system?
     – What experiences does the policy generate for the various policy actors?
     – How do staff, clients and resources interact?
     – How do beliefs and actions link to outcomes?
     – How do the actors account for the policy outcomes?
     – What are the strengths and weaknesses of the policy that are identified by policy actors, and what is the nature and degree of consensus about how well the policy works?

It is important to recognize that the much-lauded *synthesis* that can result from a mix of summative and formative evaluation does not happen by accident. It has to be managed since a number of factors conspire to inhibit it, including:

- The narrow methodological training of most researchers;
- The methods-bound perspectives of academic disciplines;
- The sheer range of skills required;
- The different time scales of quantitative and qualitative research;
- The structure of research projects that makes integration in the design analysis and interpretation difficult;
- A lack of good team management that fails to deliver easy communication, openness, trust, time and support (Walker, 2008; Bryman, 2007).

- *Theories of change:* This is arguably a unique model of evaluation in its own right but it is included here because of the profound influence that it is beginning to have on all kinds of evaluation, summative and formative (Anderson, 2005; Chen, 1990; Griggs et al., 2008; Rogers et al., 2000).

  The distinguishing characteristic of the theories of change strategy is to seek to understand people's perception of the sequence of causation thought to link policy inputs, implementation and outcomes. When and how this is done varies according to the evaluation goals:

  - It is sometimes used (as advocated above) to elicit a causal model from the policy designers early in the evaluative process, and to design the evaluation methodology to test or establish the validity of their theory of change, that is, their understanding of the way the policy works.
  - It is possible seek and compare the causal models of different groups of policy actors in order to investigate whether a lack of congruity in their perceptions helps to account for policy failings.
  - In a participatory fashion, analogous to action research, evaluators may engage with policy actors to devise or refine causal models with a view to developing and improving policy design and implementation.

- *An alternative to impact evaluation:* Formative evaluation is also used in situations when quantitative summative evaluation is impossible or inappropriate. Patton (2002) refers to "state of the art" situations when no acceptable, valid or reliable measures exist. At a more mundane level, it may also be used when, for practical reasons of time, resources or lack of forethought, a thorough impact evaluation was never put in place, or because there are indications that a policy is failing and this needs to be understood and analysed.

2. *Realistic evaluation*

   Realistic evaluation rejects experimental and quasi-experimental methods on practical and philosophical grounds. It claims that results of experiments have been found not to be generalizable because policies work differently in different contexts. Policy experiments do not have "external validity".

   This practical criticism is accompanied by a philosophical one. Unlike laboratory experimentation in which scientists go to great lengths to control the conditions in which experiments are conducted, realist evaluators would contend that policy experimentation presumes "constant conjunction", that is, that "like will always produce like".

In contrast, realist evaluation "understands causality in terms of underlying causal mechanisms generating regularities" (Tilley, 2000, p. 5). It seeks to understand the conditions under which causal mechanisms are activated to produce specific outcomes. The intended product of a realistic evaluation is therefore, to use the jargon, a "context mechanism outcome configuration" (CMOC). In order to achieve this understanding, realistic evaluators ask a set of questions about the policy intervention under consideration (Tilley, 2000, p. 7).

- *Mechanism*: What is it about a measure which may lead it to have a particular outcome pattern in a given context?
- *Context:* What conditions are needed for a measure to trigger mechanisms to produce particular outcome patterns?
- *Outcome pattern:* What are the practical effects produced by causal mechanisms being triggered in a given context?
- *CMOCs:* How are changes in outcomes produced by measures introduced to modify the context and balance of mechanisms triggered?

Unlike impact evaluation, realistic evaluation does not prioritize a specific design but, instead, typically follows a basic strategy within which the evaluator employs and adopts a range of empirical methods, both qualitative and quantitative, to facilitate measurement, comparison and understanding. The strategy approximates to the following sequence of repeated comparisons in which hypotheses about the relationship between context, mechanisms and outcomes are tested against regularities observed in existing or new instances of the intervention to provide the basis for an explanatory description:

- The process starts with theorizing about the mechanisms by which an intervention in pre-existing contexts can generate outcomes.
- This includes theoretical analysis of policy mechanisms, contexts and expected outcomes often based on analogy and metaphor.
- Hypotheses are then derived from the theoretical analysis addressing questions such as:
  - What will be the outcomes or changes brought about by the intervention?
  - What features of context will help shape these outcomes?
  - What mechanisms (social, cultural and administrative) would foster or inhibit outcomes?
- Appropriate data collection methods are then devised to investigate these processes and to provide evidence that the intervention does or does not change reality in the way hypothesized.
- On the basis of this evidence, the theory may be developed with new hypotheses and the intervention may be refined.
- Further empirical evidence will be gathered until the theory is well populated with evidence and appears robust.

Realist evaluation is best thought of as "theory driven" or "theory of change" – evaluation but distinguished by the fact that the constituents of the theories are specified in realist terms. In practice, this means that responsibility for the evolving theory of change is lodged with the evaluator and that the theory is accorded the status of "scientific", a product of repeated testing, in contrast to being "vernacular", grounded in the experiential knowledge of policy actors.

Evaluators committed to experimentation are likely to argue that evidence underpinning a theory of change is supportive rather than definitive since the general absence of a counterfactual undermines measurement of true outcome. They might also point to meta-evaluation: systematic quantitative analysis of large numbers of randomized experiments which explores the effect of context and implementation on the impact of policies (Bloom et al., 2001; Cebulla et al., 2005).

### 8.7.2 Monitoring and retrospective evaluation

Monitoring and retrospective evaluation – addressing the evaluative questions "Is policy working?" or "Did policy work?" – form part of the normal process of policy audit. Retrospective evaluation may also be triggered by the suspicion, often aroused by monitoring, that the policy is not working well.

These modes of evaluation do not require the same level of institutional commitment to evidence based policy-making as programme evaluation does. This is because:

- They are not, for example, located on the critical path from policy idea to policy introduction;
- They are generally cheaper than programme evaluation, but tend to answer different questions in different ways.

Asking evaluative questions of the present or past makes the use of counterfactuals difficult and hence unusual. This means that retrospective evaluation and monitoring pay more attention to establishing (Walker, 2004):

- Resource inputs and their conversion into service provision (*administrative efficiency*);
- The contribution of service delivery to meeting policy objectives (*administrative effectiveness*);
- The proportion of the eligible population that actually receives the social security benefit (*take up*); and
- The proportion of social security benefit recipients who wrongly receive payments because of poor policy design, maladministration or fraud (*targeting efficiency*).

The methodologies employed in retrospective evaluations tend to be eclectic and adverse circumstances can stimulate creative designs:

- Pluralistic approaches are often used in which the experiences and opinions of key actors in the policy implementation are collated and *triangulated* to reach an overall judgement on policy effectiveness;
- Personal interview surveys may be used to solicit the views of policy recipients;
- Qualitative interviews conducted with administrators and other interest groups and observational techniques may be used at the point of service delivery;
- The accumulation of accounts obtained in these ways in mixed methods studies can provide irrefutable evidence about the efficiency or lack of efficiency of implementation and provide a sound basis for reform.

Where retrospective evaluations have sought to assess policy impact they have typically used one or more of three approaches:

- *Trend analysis*, which involves inspecting time series data on the outcome indicator to identify an inflection coinciding with the introduction of the policy. More sophisticated analyses use time-series regression or other simulation techniques to define a counterfactual by predicting the trend of the variable in the absence of the policy and comparing the prediction with the actual trend (White and Riley, 2002).
- *Difference in difference techniques*, which is used when the experiences of the group targeted by the policy can be compared with an ostensibly similar group who do not receive it (Hasluck et al., 2000).
- *Judgemental assessment*, in which policy actors, administrators, benefits recipients and/or engaged professionals are asked directly to assess the effectiveness of a policy on the basis of their own experience, or with respect to cases that they have processed, and their assessments collated and perhaps triangulated (Thornton and Corden, 2002).

## 8.8 Ethics of policy evaluation

While policy evaluation falls under the same ethical rubric as policy research, prospective evaluation raises the threshold of necessary vigilance:

- Prospective forms of evaluation (e.g. programme evaluations, piloting and prototyping) involve establishing and implementing social security schemes on a small scale to test whether the policy idea could work in practice;
- Participants in such evaluations are therefore explicitly placed in situations where the long-term outcomes for individuals are uncertain;
- This adds to the risks associated with the research process, those connected to the consequences of the policy intervention itself:
  - Persons allocated to the *control group* will be denied any positive benefits arising from the social security scheme being evaluated;
  - Persons allocated to the *programme group* may suffer a loss in well-being if the experimental programme proves not to work.
- The justification used for imposing pilot programmes in spite of these additional risks is that the resultant policy knowledge is going to be used to enhance the common good.

The principles underpinning evaluation research can be reduced to three: beneficence, justice and respect for persons (Blustein, 2005).

### 8.8.1 Beneficence

The beneficence principle states that harm should not be done to research subjects. However, given that this cannot be guaranteed in a policy experiment, the requirement is that potential harm should be minimized and possible benefits maximized.

The key ethical test is uncertainty as expressed in the *sub-principle of equipoise*:

- It is essential that there is uncertainty as to whether the experimental programme works and hence whether people in the programme or control group will fare better, otherwise one group will be knowingly disadvantaged.
- Therefore if early results are conclusive, the policy experiment should arguably be curtailed as often happens in medical drug trials.

In practice, policy trials tend to offer additional presumed "benefits", such as cash payments and additional advice. It is therefore assumed that the risk of "harm" to the trial group is not as great as in medical trials where the impact of the intervention is unknown.

The beneficence principle is also generally interpreted to mean that existing provision cannot be withdrawn to create a control group. This means that the overall efficacy of a new programme cannot be directly established, only its performance relative to the existing one.

### 8.8.2 Justice

The principle of justice seeks to strike the balance between the interests of the research subjects and benefits to society (Blustein, 2005).

This balance is often difficult to divine. Evidence cited of injustice include:

- Social security recipients having less political leverage than taxpayers;
- The tendency for public policies to be piloted on less politically powerful groups, welfare recipients rather than the middle class.

Equally, low status individuals are more reliant on public services than higher status groups and are therefore disproportionately likely to be involved in a policy trial.

Issues to do with justice have also arisen in debates about the legitimacy of individual random assignment. Some critics argue that random assignment is unjust on the grounds that:

- Even if the policy being tested involves the transfer of resources this cannot be presumed to be individually advantageous if the outcome of the policy is truly unknown;
- It is unfair to allocate a public good on the basis of chance rather than need.

Others defend the use of random assignment on grounds of justice opining that:

- If resources are limited and it is not known who might benefit most, then chance is as equitable a method of resource allocation as any.

Even so, these arguments are not necessarily sufficient to convince everyone that random assignment is appropriate. In Britain, for example, random assignment is employed much less frequently than geographical controls because politicians seem more comfortable to differentiate between similar individuals living in different areas than similar ones living as neighbours (Cabinet Office, 2004; Walker and Duncan, 2007). There may be no clear moral basis for this distinction but the political logic is that it reduces the chances of voters in the same constituency being treated differently in a policy experiment.

### 8.8.3 Respect

The principle of respect for persons "*holds that subjects should be treated as autonomous agents, with their own perspectives, goals, values and considered opinions*" (Blustein, 2005, p. 826).

From this flow the requirements that:

- Participants should give their consent to participate;
- Consent should be informed, since only the individuals themselves know whether they wish to participate in an evaluation.

This, in turn, requires two conditions to be fulfilled (Orr, 1999):

- Participants are given sufficient information – and that they understand it – to allow them meaningfully to choose participating and not participating;
- Consent is obtained fairly and without coercion.

Obtaining informed consent is not always easy and there is increasing evidence that these principles are frequently breached, often inadvertently and sometimes deliberately.

- Participants in policy experiments may understand neither the policy nor the evaluation, suggesting that any consent which was secured was not informed (Stratford et al., 2005; Walker et al., 2008; Wiles et al., 2005).
- Participants involved in medical trials have been reported to underestimate the chances of being harmed and to overstate the likely efficacy of new drugs (Joffe et al., 2001; Lidz et al., 2004).
- People living in control areas are seldom told about the social experiment in which they are nominally involved (inaction that may be defensible in that they are not directly affected).
- Informed consent may compromise the design of the evaluation as when, for example, alerting someone to the possibility of receiving training might encourage them to seek it out.

## REFERENCES


Anderson, A. 2005. *The community builder's approach to theory of change: A practical guide to theory and development*. New York, NY, Aspen Institute Roundtable on Community Change.

Bloom, H.; Hill, C.; Riccio, J. 2001. *Modelling the performance of welfare-to-work programs: The effects of program management and services, economic environment, and client characteristics.* New York, NY, Manpower Demonstration Research Corporation.

Blustein, J. 2005. "Toward a more public discussion of the ethics of federal social program evaluation", in *Journal of Policy Analysis and Management*, Vol. 24, No. 4.

Bottomley, D.; Walker, R. 1996. *Experimental methods for policy evaluation* (Working paper, No. 276S). Loughborough, Loughborough University – Centre for Research in Social Policy.

Bryman, A. 2007. "Barriers to integrating quantitative and qualitative research", in *Journal of Mixed Methods Research*, Vol. 1, No. 1.

Burtless, G.; Orr, L. 1986. "Are classical experiments needed for manpower policy?", in *Journal of Human Resources*, Vol. 21, No. 4.

Cabinet Office. 1999. *Professional policy-making for the twenty-first century* (Report by the Strategic Policy-Making Team). London.

––. 2004. *"Trying it out": Review of the effectiveness of government pilots.* London.

Case, A.; Deaton, A. 1998. "Large cash transfers to the elderly in South Africa", in *The Economic Journal*, Vol. 108, No. 450.

Cebulla, A.; Ashworth, K.; Greenberg, D.; Walker, R. 2005. *Welfare to work: New Labour and the US experience.* Aldershot, Ashgate.

Chen, T. 1990. *Theory driven evaluation.* New York, NY, Sage.

Federal Register. 1991. "The federal policy for the protection of human subjects", in *Federal Register,* Vol. 56, No. 117.

Galasso, E.; Ravallion, M.; Salvia, A. 2001. *Assisting the transition from workfare to work: A randomized experiment* (Policy research working paper, No. 2738). Washington, DC, World Bank.

Griggs, J.; Whitworth, A.; Walker, R.; McLennan, D.; Noble, M. 2008. *Person or place-based polices to tackle disadvantage? Not knowing what works.* York, Joseph Rowntree Foundation.

Hasluck, C.; McKnight, A.; Elias, P. 2000. *Evaluation of the new deal for lone parents: Early lessons from the phase one prototype – cost-benefit and econometric analyses* (DSS research report, No. 110). London, Department of Social Security.

JMMR. 2007. "The new era of mixed methods: Editorial", in *Journal of Mixed Methods Research*, Vol. 1, No. 1.

Joffe, S.; Cook, E. F.; Cleary, P. D.; Clark, J. W.; Weeks, J. C. 2001. "Quality of informed consent in cancer clinical trials: A cross-sectional survey", in *The Lancet*, No. 358.

Karlan, D.; Valdivia, M. 2006. *Teaching entrepreneurship: Impact of business training on microfinance clients and institutions* (Working paper, No. 941). New Haven, CT, Yale University Economic Growth Center. <http://aida.econ.yale.edu/karlan/papers/TeachingEntrepeneurship.pdf>

––; Harigaya, T.; Nadel, S. 2006. *Evaluating microfinance program innovation with randomized controlled trials: Examples from business training and group versus individual liability* (Conference paper, 2006 Global Microcredit Summit, Halifax, 12-16 November).
<http://library.financialaccess.org/pdf/I15_FAI_TeachingEntrepeneurship.pdf>

Kazi, M. 2003. *Realist evaluation in practice: Health and social work.* London, Sage.

King, G. 2007. "A 'politically robust' experimental design for public policy evaluation, with application to the Mexican Universal Health Insurance Program", in *Journal of Policy Analysis and Management*, Vol. 26, No. 3.

Lidz, C. W.; Appelbaum, P. S.; Grissoa, T.; Renaud, M. 2004. "Therapeutic misconception and the appreciation of risks in clinical trials", in *Social Science & Medicine*, Vol. 58, No. 9.

Morton, M. 2008. *Applicability of impact evaluation to cohesion policy* (Unpublished mimeo). Oxford, University of Oxford – Department for Social Policy and Social Work.

Orr, L. 1999. *Social experiments: Evaluating public programs with experimental methods* Thousand Oaks, CA, Sage.

Patton, M. 2002. *Qualitative research and evaluation methods.* Thousand Oaks, CA, Sage.

Pawson, R.; Tilley, N. 1997. *Realistic evaluation.* London, Sage.

Rawlings, L.; Rubio, G. 2003. *Evaluating the impact of conditional cash transfer programs: Lessons from Latin America* (Policy research working paper, No. 3119). Washington, DC, World Bank.

Rodriguez-Planas, N. 2007. *What works best for getting the unemployed back to work: Employment services or small-business assistance programmes? Evidence from Romania* (IZA discussion paper, No. 3051). Bonn, Institute for the Study of Labor.

Rogers, P.; Hacsi, H.; Petrosino, A.; Huebner, T. 2000. *Program theory in evaluation: Challenges and opportunities* (New directions for evaluation series, No. 87). San Francisco, CA, Jossey-Bass.

Stratford, N.; Farrell, C.; Natarajan, L.; Lewis, J. 2005. *Taking part in a randomised control trial: A participant's eye-view of the Job Retention and Rehabilitation Pilot* (Research report, No. 273). London, Department for Work and Pensions.

Thornton, P.; Corden, A. 2002. *Evaluating the impact of access to work: A case study approach* (Research & development report, No. WAE138). Sheffield, Department for Work and Pensions – Claimant Unemployment and Disadvantage Analysis Division.

Tilley, N. 2000. *Realistic evaluation: An overview* (Conference paper, Founding Conference of the Danish Evaluation Society, Aarhus, September). <http://www.evidence-basedmanagement.com/research_practice/articles/nick_tilley.pdf> (accessed on 31.10.2010).

Walker, R. 2000. "Learning if policy will work: The case of New Deal for Disabled People", in *Policy Studies*, Vol. 21, No. 4.

––. 2008. "Mixed methods research: Quantity plus quality", in J. Neale and A. Root (eds.), *Research methods for health and social care.* Basingstoke, Palgrave Macmillan.

——; Duncan, S. 2007. "Knowing what works: Policy evaluation in central government", in H. Bochel and S. Duncan (eds.), *Making policy in theory and practice.* Bristol, Policy Press.

——; Hoggart, L.; Hamilton, G. 2008. "Random assignment and informed consent: A case study of multiple perspectives", in *American Journal of Evaluation*, Vol. 29, No. 2.

Walker, R. 2004. "Evaluation: Evidence for Public Policy", in OECD (ed.): *Evaluating Local economic and employment development: How to assess what works among programmes and policies*. Paris, Organisation for Economic Co-operation and Development.

White, M.; Riley, R. 2002. *Findings from the macro evaluation of the New Deal for Young People*. London, Department for Work and Pensions.

Wiles, R.; Heath, S.; Crow, G.; Charles, V. 2005. *Informed consent in social research: A literature review* (NCRM methods review papers, No. 001). Southampton, ESRC National Centre for Research Methods.

Yap, Y.-T.; Sedlacek, G.; Orazem, P. 2002. *Limiting child labor through behavior-based income transfers: An experimental evaluation of the PETI program in rural Brazil.* Washington, DC, World Bank.

Promoting and developing social security worldwide
Promouvoir et développer la sécurité sociale à travers le monde
Promover y desarrollar la seguridad social en el mundo
Soziale Sicherheit weltweit fördern und entwickeln
Развиваем и поддерживаем социальное обеспечение во всем мире
دعم و تطوير الضمان الإجتماعي عبر العالم
促进和发展全球社会保障

4 route des Morillons
Case postale 1
CH-1211 Geneva 22

T: +41 22 799 66 17
F: +41 22 799 85 09
E: issacomm@ilo.org | www.issa.int