NHS

Health Development Agency

Evaluation of community-level interventions for health improvement: a review of experience in the UK

Dione Hills

Tavistock Institute

This document is also published on the Health Development Agency website at www.hda.nhs.uk

Copies of this publication are available to download from the HDA website (www.hda.nhs.uk).

Health Development Agency
Holborn Gate
330 High Holborn
London
WC1V 7BA

Email: communications@hda-online.org.uk

ISBN 1-84279-303-9

© Health Development Agency 2004

About the Health Development Agency

The Health Development Agency (www.hda.nhs.uk) is the national authority and information resource on what works to improve people's health and reduce health inequalities in England. It gathers evidence and produces advice for policy makers, professionals and practitioners, working alongside them to get evidence into practice.

Contents

Summary **1**
Aims 1
Overview 1
The research challenge 1
Common evaluation approaches 2
Evaluation quality 2
Need for a match between programmes, evaluation questions and evaluation design 3
1 Mapping the territory **4**
Introduction to the task and rationale for the work 4
Methodology 4
The challenge of a systematic literature review 5
Breadth of the area and difficulties in establishing boundaries 5
Characteristics of community-level interventions 6
Changing conceptual understanding 8
Defining evaluation 10
Difficulty in accessing published accounts 11
Criteria adopted in selecting literature 12
Overview 12
2 Contextual factors affecting the evaluation of community health initiatives **13**
Introduction 13
Community health interventions pre-1997 13
Community health initiatives post-1997 15
Changing expectations of evaluation 15
Conclusions 17
3 The evaluation challenge **18**
Introduction 18
Challenging nature of the interventions 18
Common evaluation strategies 20
Strengths and weaknesses of these approaches 24
Conclusions 25
4 New approaches to resolving evaluation dilemmas **26**
Introduction 26
Tackling context and theory 26
Frameworks for programme evaluation 28
Dealing with programme complexity 29
New thinking about outcomes and indicators 31
Meta-evaluation and systematic reviews 33
Conclusions 34

5 The changing evaluation context – implications for good practice **36**
Introduction 36
Establishing quality criteria 36
Evaluation utility 37
Feasibility and appropriateness of the evaluation strategy 38
Propriety in evaluation practice – the centrality of ethics and values 40
Methodological competence and accuracy in reporting 42
Good practice in commissioning and dissemination 43
Gathering evidence of good evaluation practice 45
Conclusions 46
6 Improving the evidence base for community health interventions **48**
Introduction 48
A good map of the territory 48
Asking the right evaluation questions 49
Building an evidence base in the field 49
Ensuring high-quality evaluations 50
References **51**
Appendix: Expert seminar: evaluating community-level interventions: where are we now? **55**

Acknowledgements

I would like to thank the following people for their support and assistance with this review:

- Vicky Blackburn, of the Tavistock Institute, whose hard work in undertaking the literature searches, and helping to make sense of the material was crucial in the early stages of the review
- Elliot Stern, of the Tavistock Institute, who contributed invaluable expertise on evaluation models and quality assurance processes
- Julie Bull, previously of the Health Development Agency, who commissioned the review, helped determine its parameters and organised the expert seminar
- Professor Mike Kelly and Antony Morgan of the Health Development Agency, who chaired the expert seminar and secured publication of the review.
- The following evaluation experts, who attended the expert seminar, several of whom also contributed valuable feedback on the final report:
  - Dr Vicky Cattel: University College London
  - Gill Cowburn: University of Oxford
  - Paula Gaunt-Richardson: Glasgow Caledonian University
  - Professor Ken Judge: University of Glasgow
  - Dr Judith Monks: University of Keele
  - Professor Pauline Ong: University of Keele
  - Dr Ray Pawson: University of Leeds
  - Professor Jenny Popay: University of Leeds
  - Mai Stafford: University College London
  - Sylvia Tilford: London Metropolitan University
  - Dr David Woodhead: The King's Fund
- And finally, to the copy-editors who helped to turn the final report into a quality production.

Summary

Aims

This literature review of the evaluation of community-level health interventions was commissioned by the Health Development Agency (HDA) to:

- Bring together current knowledge about different approaches to the evaluation of community-based programmes
- Identify recent innovations and promising examples
- Begin to identify what constitutes quality in the principles of evaluation design, so as to inform and encourage robust contributions to the emerging evidence base.

The review focuses on recent British literature and examples incorporating innovative work from abroad where relevant. It was undertaken via a search of key sources, including electronic databases, journals and websites. The review also includes additional material gained from an HDA Expert Seminar held in April 2001 (see Appendix), particularly concerning issues of evaluation quality.

Overview

Chapter 1 provides an introduction to the review, including a summary of current debates surrounding community-level health interventions and their evaluation.

Chapter 2 describes the context in which evaluations in this area have taken place. This is crucial background for understanding the evaluation approaches that have been adopted at various points over the past 15 years in the UK.

Chapter 3 explores some of the main challenges facing evaluators working in this area. It reviews the main evaluation strategies that have been used until recently in terms of their strengths and weaknesses.

Chapter 4 examines some of the newer evaluation strategies that have emerged over the past six years, particularly since the establishment of a new range of large national health programmes incorporating community participation.

Chapter 5 provides an overview of some of the issues related to quality in evaluation.

Finally, Chapter 6 sets out some of the steps that need to be taken to improve the quality of evaluation in the future.

The research challenge

The task of undertaking a review of this nature was challenging because of:

- Lack of publications and difficulty in accessing evaluation reports
- Breadth – difficulties in setting a clear boundary around the area
- Changing conceptual understandings in the area.

Community-based interventions for health improvement are not a discrete set of activities, but are a range of very varied initiatives, programmes, projects and activities emerging out of a combination of the following areas:

- Health promotion
- Community development
- Regeneration/area-based initiatives
- Moves towards greater participation in health service planning
- Moves towards promoting ‘active citizens’ and ‘active patients’
- Debates about social capital
- Health partnerships/health alliance activities
- Community arts activities
- Attempts to address health inequalities.

To put a boundary around the area, various models are put forward which help clarify what constitutes a community-level or health intervention.

To aid understanding of the kinds of evaluation that have been taking place in this field, a broad review is presented of some of the contextual factors that have influenced decisions taken:

- Nature of the community health initiatives taking place
- Wider political and professional context
- General trends in evaluation theory and practice.

Until recently, health interventions at community level in the UK have been relatively small scale and local, with evaluation activities often driven by the need to obtain and retain funding. A lack of policy interest and a lack of funding have meant there was little pressure to undertaken evaluations from which lessons for wider policy and practice could be learned.

However, a growing interest in community-based interventions and the commissioning of a number of large-scale programmes since 1997 have stimulated a demand for large-scale evaluations, and provided new opportunities for exploring evaluation approaches not widely used previously in the UK. The new culture of accountability has challenged the evaluation community to come up with new approaches to evaluation and outcome measurement.

Common evaluation approaches

During the early 1990s (pre-1997), evaluation of community-level interventions was largely influenced by three different traditions:

- **Experimental** and **quasi-experimental** designs
- **Participatory methods** including empowerment, stakeholder, participatory and illuminative research strategies
- **Multi-method strategies** that bring together a variety of research approaches, usually incorporating considerable use of qualitative data and process indicators.

Pluralistic evaluation and participatory and empowering evaluation strategies have been widely used in community-level interventions as a means of strengthening programme development, but often fail to generate generalisable results. Experimental and quasi-experimental approaches have occasionally been used, can be difficult to implement, and may be unpopular with programme participants.

Newer evaluation approaches

The shortcomings of these approaches have been addressed by the development of new evaluation approaches, including:

- **Theory-based** and **realistic** evaluation strategies: these challenge the lack of attention to context, programme theory and the mechanisms of change in earlier experimental and quasi-experimental approaches
- **New evaluation frameworks**: these build on earlier multi-method approaches by providing frameworks that indicate different evaluation strategies at various stages, and at different levels, in a programme or project
- New thinking about the **nature of complexity** and adoption of **socio-ecological theoretical frameworks**: these are still at an early stage of development in terms of evaluation, and have yet to be developed into fully operational evaluation strategies
- Fresh thinking in the area of **indicators and outcome measures**: provides new ways of evaluating the structural changes that arise as a result of community-based projects
- Development of **new approaches to systematic reviews** that are less dependent on experimental and quasi-experimental designs.

Evaluation quality

Assessment of evaluation needs to take account of a number of factors beyond the appropriateness of the methodology, and includes consideration of utility, feasibility, propriety and accuracy. The processes for commissioning, collating and dissemination of evaluation findings also need to be taken into account.

Utility

Clarifying the different purposes for which evaluation might be undertaken – for development of the project, to generate generalisable knowledge, or for accountability purposes – can help ensure evaluation strategies are appropriate to their purpose.

Feasibility

Expectations of evaluation can often outstrip what is feasible, given the resources available and the capacity of a given intervention to sustain or deliver. There is a need for realism among those commissioning and undertaking evaluations regarding what can be expected of a local project and the community in which it takes place.

Propriety

Evaluation strategies need to be broadly consistent with the values and aims of the intervention itself. In the field of community health interventions, this suggests evaluation strategies that have the capacity to involve and empower community members. Values such as courtesy, openness and the appreciation of diversity are particularly important in evaluation strategies that require close engagement between evaluators and community members.

Methodological competence, accuracy and transparency

Lack of crucial information about methods, the nature of an intervention and its context, and the level of involvement of community members makes it difficult to judge the accuracy of methods and their appropriateness to the intervention taking place. Essential information required for replication, or for drawing generalisable conclusions about the efficacy of the intervention, is often left out.

Given the wide use of qualitative and participatory evaluation methods in this field, the availability of new, structured approaches to the assessment of quality in this kind of research is welcomed.

Commissioning and dissemination

The quality and utility of an evaluation is strongly influenced by the commissioning process, and by how evaluation results are collated and disseminated. Work from the European Union (EU) suggests that quality assurance mechanisms can be built in at all stages in the evaluation process – something that could usefully be applied to community-level health interventions.

Need for a match between programmes, evaluation questions and evaluation design

The review of evaluation strategies suggests that considerable advances have been made in recent years in finding new ways of combining both rigorous quantitative and qualitative and participatory approaches. However, evaluators and commissioners in this field are still experiencing difficulties, which can be compounded by the:

- Range of different conceptualisations or discourses operating in the area, concerning both the practice of community-level work and evaluation
- Lack of clear mapping of the territory, including the nature of community-level interventions, types of evaluation, and the different evaluation questions that need to be addressed
- Mismatches between the kind of evaluation questions being asked, the nature of the interventions of which they are being asked, and the types of evaluation strategies adopted.

This review contributes towards mapping the various evaluation strategies that have been adopted over the past 15 years, but there remains a lack of theoretical mapping of the kinds of community health interventions taking place and the evaluation questions that need to be answered. While this gap may be filled in from the evaluations of some of programmes currently being funded (healthy living centres, health action zones, Sure Start, Social Action Research Project), this will be useful only if steps are taken to collate and disseminate this experience to those planning new policies and programmes in this field.

1 Mapping the territory

Introduction to the task and rationale for the work

This literature review was commissioned by the Health Development Agency (HDA) because of its concern about an 'evaluation deficit' in community-based interventions for health improvements. Although community-level approaches to health improvement have shown promise for many years now, the evidence base for this area of work was believed to be problematic and underdeveloped.

The renewed emphasis on area- and community-based interventions in current policy, and the funding of major initiatives such as Sure Start, health action zones (HAZs) and healthy living centres (HLCs), has stimulated the evaluation community to explore new models and approaches for evaluating such initiatives. The aim of this review was to capture both past and present learning by:

- Bringing together current knowledge about different approaches to the evaluation of community-based programmes for dissemination to the research/evaluation community, and to practitioners concerned with evaluation issues
- Identifying recent innovations and promising examples in order to share learning
- Beginning to identify what constitutes quality in the principles of evaluation design, leading to robust contributions to the evidence base.

The review focuses on recent UK work, although reference is made to examples and debates from abroad where relevant. On the whole, the amount of work taking place in the UK has been on a fairly small scale compared with the large programmes that have been evaluated elsewhere. Those wishing to explore this topic in depth are advised to read this review alongside the much larger international examination of evaluation approaches used in the field of health promotion by the World Health Organization (WHO) Europe (Rootman *et al.*, 2001).

Methodology

The review was undertaken initially by searching for articles and reports through electronic databases and manual searches of relevant journals, using the keywords 'community level', 'community participation' and 'community development', together with 'evaluation' and 'evaluation review'. Although this threw up many articles discussing the nature of this kind of work and the difficulties of evaluating it, the search revealed relatively few examples of actual evaluations. In order to find these, it was necessary to search out grey literature through an examination of websites and discussions with people working in agencies specifically involved in work of this kind.

Box 1.1 Key sources for the review

Electronic databases of abstracts, including the International Bibliography of the Social Sciences (IBSS), MEDLINE, the Thomson/ISI Web of Science, the HDA's HealthProm*is* database and the King's Fund's library database

Specific searches in the journals *Evaluation*, *Evaluation Review*, *Community Development Journal*, *Community Health UK*, *ARVAC Bulletin*

Web searches of sites related to evaluation, community development, public health and community health

Search of grey literature including evaluation conference abstracts, unpublished reports and papers pending publication

Review of bibliographies from key papers.

Box 1.2 Characteristics of community-based versus community development models (after Labonte, 1992)

Community-based	**Community development**
Problem, targets and action defined by sponsoring body	Problem, targets and action defined by community
Community seen as medium, venue or setting for intervention	Community itself the target of intervention in respect to capacity building and empowerment
Notion of 'community' relatively unproblematic	Community recognised as complex, changing, subject to power imbalances and conflict
Target largely individuals within either geographic area or specific subgroup in geographic area defined by sponsoring body	Target may be community structures or services and policies that impact on the health of the community
Activities largely health-oriented	Activities may be quite broad-based, targeting wider factors with an impact on health, but with indirect health outcomes (empowerment, social capital)

A selection was then made of the reports and articles using the criteria identified on page 12. These were examined to see what evaluation strategies had been used, and the advantages or disadvantages of different approaches identified. Additional material from policy documents, historical accounts and the wider evaluation literature was added to provide contextual understanding of why different approaches to evaluation were being used at different times.

Some initial findings were debated by a seminar of evaluation experts (see Appendix). This also provided a forum for discussion of good practice in relation to evaluating community-based initiatives. The proceedings of this Expert Seminar were recorded, and individuals' observations have been used to illustrate points that had also emerged from the literature review, particularly in Chapter 5.

The challenge of a systematic literature review

Although initially envisaged as a relatively limited undertaking, compiling a systematic overview of the literature in this area proved surprisingly difficult. The reasons for this give a clue to the difficulties in using evaluations of such activities to develop a systematic evidence base. The difficulties fall into five categories:

- Breadth of the area and difficulties in establishing a boundary around community-level interventions as a discrete area of work
- Characteristics of community-level interventions
- Nature of evaluation in this area
- Difficulty in accessing published accounts of evaluations
- Changing conceptual world in which this work takes place.

Breadth of the area and difficulties in establishing boundaries

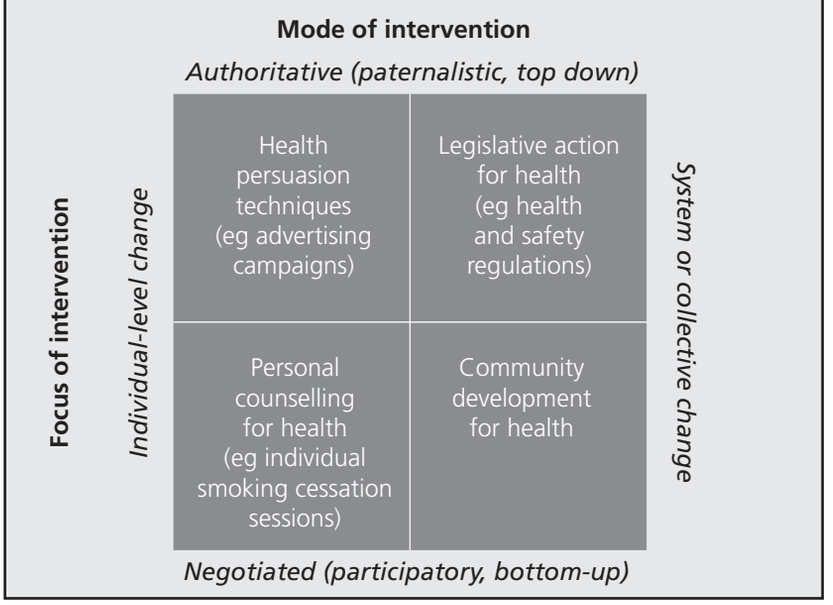
The initial starting point was to undertake a review of different approaches to the evaluation of community-based interventions for health improvement. However, it was soon apparent that these do not exist as a discrete set of activities: at best, community interventions for health are a range of very varied initiatives, programmes, projects and activities emerging out of, and overlapping with, the following areas:

- Health promotion
- Public health
- Community development
- Activities addressing health inequalities
- Regeneration/area-based initiatives
- Moves towards greater participation in health service planning
- Moves towards promoting 'active citizens' and 'active patients'
- Attempts to increase 'social capital' and to enhance 'community capacity'
- Health partnerships/health alliance activities
- Community arts/arts/music for health.

The situation is particularly complex on the ground at a time when many programmes are being established to address social inequalities, particularly health inequalities. In many deprived areas, a number of different programmes operate alongside one another, often with similar and overlapping aims. Distinguishing those with specific health aims from those with more general employment or environmental improvement aims is often far from easy. At a time when

Box 1.3 Different approaches to health promotion

(Source: Beattie, 1991b)



involvement of the community has become central to many government initiatives, distinguishing a specifically community-level intervention from other health initiatives can also be difficult.

Characteristics of community-level interventions

The major difficulty in putting a clear boundary around community-level health interventions lies in the heterogeneous and complex nature of such activities. Various frameworks have been used to distinguish such interventions from other health or broader area-based interventions.

Community-based versus community development activities

In public health initiatives there is an important conceptual difference between health promotion activities which are provided in the community, and those which actively seek to involve people from the community, either in setting the agenda or in delivering the activity. Labonte (1992) uses the terms 'community-based' and 'community development' to make this distinction (Box 1.2).

System-wide change

This distinction is similar to one put forward by Beattie (1991b) in his analysis of different types of health promotion taking place in the 1980s (Box 1.3).

The key difference here is not only the extent to which the community is taking an active role in setting the agenda, but also the fact that the target of change is not just the individual, but the social, cultural and organisational environment which influences individual behaviour. In many cases, community-based initiatives are as much about establishing an infrastructure – of community groups, partnerships between agencies and new, cross-cutting structures – as they are about the delivery of specific health-related activities to individuals.

Rationale or value base

The change of emphasis from top-down provision – based on a professionally determined health agenda – to community-based provision also means working according to a different set of principles or values.

Scott and Weston (1998) based their set of definitions of health promotion activities on the underlying rationale of projects and programmes, distinguishing between community development approaches and those based on biomedical, prevention or educational models of health intervention.

- **Biomedical model**: grounded in principles of medical science such as public health medicine and epidemiology
- **Prevention model**: grounded in principles of medical science but recognising that education and information programmes are necessary, and including more interactive relationship with patients, individuals and communities
- **Education model**: grounded in educational principles, based on sharing of knowledge and information with communities, also concerned with issues of self-esteem, autonomy and empowerment

- **Community development model**: integrates prevention and education models, but seeks to empower communities, recognising differences in culture, economic and political realities
- **Radical or political model**: seeks change through health policy development, addressing issues of equity, justice, human rights and 'healthy public policy'.

The underlying rationale is important in two respects. First, it suggests that similar activities may sometimes be undertaken for very different reasons – incorporating a different theory concerning the link between the activity and its intended impact. Second, Scott and Weston (1998) also found that it influenced the choice of evaluation strategy adopted. For example, evaluations of interventions based on the medical model are likely to be grounded in an empiricist research tradition, while educational or community development interventions tend to adopt a qualitative research tradition, and radical interventions tend towards a radical/political research tradition.

Multiple activities

The difference between activities *in* the community and activities undertaken *with* the community is also central to a distinction made by Green and Kreuter (1991) between large-scale, community-wide programmes working *with* the community, and single-focus health promotion projects taking place *in* it. The former generally seek relatively small but pervasive changes for most or all of the population, while the latter seek intensive and profound changes in the risk behaviour of a small subsection. Potvin and Richard (2001) make a similar distinction between community programmes and small-scale, community-based activities, the former being large, complex, participatory, long-lived, flexible and adaptable.

A central distinction here is that many community-level programmes are not simply delivering one intervention with a specific goal, but are responding to local health concerns by establishing a programme of activities, adapting these to changing circumstances, and responding to opportunities that arise from the community. Each programme, even when part of a national initiative, is therefore unique, and any attempt to standardise them would run counter to the rationale of the approach. This has important implications for any attempt to draw comparisons between programmes or to find out which particular feature accounted for its success or failure.

It also has implications for the focus of evaluation. For example, an evaluation may focus on specific activities – as in project work – or on how a number of projects and activities have been combined within a programme. In many community-level programmes it is this combination – including the way the programme built partnerships between different organisations or set about engaging the community in activities – that is at the heart of the innovation.

Box 1.4 Health – a central or peripheral concern for community-level interventions?

No specific health targets
Interventions with no specific health aims (eg programmes sponsored by Home Office Active Community Unit, crime-prevention measures, area-based interventions to combat social exclusion etc)
Interventions with improvement in community health as one aim among several (eg New Deal for Communities)
Interventions in which health improvement is a primary objective, although activities not specifically targeted at health (eg Social Action Research Project, in which building social capital is primary aim)
Interventions in which activities have a broad – and positive – health focus (eg HAZs, HLCs)
Interventions that have a very specific health focus (eg community-based health promotion programmes to tackle HIV/AIDS, smoking etc)
Very specific health targets

Broad definition of health

Most community-level interventions start from a very broad definition of health, and often place an emphasis on positive health rather than on simple prevention of illness (Davies and Kelly, 1993). This broad approach to 'health', together with the current interest in health inequalities, means that the boundary between interventions that have a health focus, and other area-based or community-oriented interventions, has become blurred. With a growing awareness of the multiple factors that affect health, many community-based interventions that seek to address other factors – unemployment, crime, poor educational standards – can also be seen to have a potential impact on health, even if indirectly.

Community-based interventions can therefore be seen to lie on a continuum between those that have only very indirect health aims, and those that have a very specific health focus (Box 1.4).

Tones *et al*. (1990) put forward similar distinctions which highlighted the fact that community-based interventions can sometimes evolve from one kind to another. Their typologies incorporate both the dimension of participation (how far the community is setting the agenda), and the extent to which the agenda relates directly or indirectly to health (Box 1.5).

Tones *et al.* (1990) identify the fact that projects can migrate from one rung to another in the course of their development. For example, funders may begin a project with a specific heart health agenda, but the community then identifies a wider set of health issues on which they wish to take action (Hills and King, 1993). Alternatively, a project may begin with a broad community development agenda, but more specific health targets emerge from discussions between local agencies and community members (as implied by Type 2, Box 1.5).

Changing conceptual understanding

Changing attitudes towards the evaluation of community-based health interventions are being affected by the wider cultural environment within which attitudes towards health, and towards community involvement, are changing. These changes affect the underlying assumptions – about appropriate interventions or appropriate models of evaluation – on which decisions are based. (Schein, 1992 argues that basic assumptions are a key factor in binding different groups or subgroups together: they are the implicit assumptions that guide behaviour and tell group members how to perceive, think and feel about things. They are taken for granted and lie largely outside the conscious awareness of group members, and as such are difficult to confront and debate.) Because such assumptions are often implicit, rather than explicit in documentation or debates relating to such activities, they can be potent sources of confusion and conflict.

In the health field there is a growing tension between, on one hand, the increasing demand for evidence-based practice and, on the other hand, growing awareness of complexity in the health field. This is particularly relevant in relation to community-based interventions with their multiple activities addressing different aspects of health, and their community-level involvement.

Box 1.5 Typologies for identifying community involvement (Tones et al., 1990)

Type 1	Innovator's goal for the community is primarily self-empowerment and improvement in socio-economic status
Type 2	As above, but in the process of developing a community profile and identifying felt needs, the community itself acknowledges needs that are consistent with standard preventive medical/health education goals, eg need for better primary care services, accident prevention, dealing with child health problems
Type 3	Characterised by community health projects. Innovator's goal is to enhance health and prevent diseases by raising profile of health, but is prepared to help community work through other more pressing 'felt needs', eg to improve cardiovascular health
Type 4	Innovator's goals are primarily those of preventive medicine. Epitomised by the various coronary heart disease prevention programmes. More 'top-down' than types 1–3, but innovator understands importance of taking the community with them and utilising existing leadership patterns
Type 5	More limited outreach programmes, limited community participation, but uses mix of agencies, eg media plus schools plus drop-in centres, and delivery of services to housing estates or workplaces

The kind of changes that are affecting thinking in the health field can be summarised as follows.

From	To
A disease prevention model targeting morbidity and mortality	A positive health approach targeting general health and wellbeing (eg the WHO model)
A model of single disease causality, eg fighting infectious epidemics	Multiple causality of health problems (as in *Our healthier nation*: DH, 1998a)
Individual level of intervention, eg health education interventions	System level of intervention – programmes such as HAZs
The notion of passive patients, eg traditional healthcare strategies	Active public/patient participation in health and health promotion, as in promotion of patient choice and expert patients initiatives
Health inequalities peripheral to health policy development	Health inequalities central to health policy development (as in *Our healthier nation*: DH, 1998a)

There are also changes taking place in how communities are being conceptualised, and incorporated into policy and practice development. A recent review of *Measures of community* undertaken for the Home Office Active Community Unit (Chanan, 2002) concluded that:

> 'A community is not a "thing". It is a number of people who have repeated dealings with each other … What must never be lost sight of, however, is that community life is voluntary and autonomous or it is nothing. Government cannot "produce" it – it can only assist it to produce itself.'

The changes in perception regarding community can be summarised as follows.

From	To
Community seen as passive setting for intervention as in previous 'community-based' services	Community as dynamic partner in health promotion activities, as in community development approaches
Community seen as relatively peripheral to policy development, as in previous approaches to regeneration of deprived areas	Community as a central resource in implementation of change, eg in many area-based initiatives such as New Deal for Communities, HAZs, Sure Start, etc
Community development seen as a means of implementing change	Community development (community capacity building) as an end in itself
Community seen as an entity	Community seen as a complex set of processes involving multiple stakeholders and shifting power relationships
Services seen as a given, and not up for negotiation (common to many earlier community-based interventions)	Services seen as potentially part of the problem being addressed and the target of change [as in the King's Fund (2000) Strategic Action Programme for Healthy Communities]

These changes have had important implications for evaluation practice, especially around issues such as the identification of the 'object' of evaluation, the choice of research method, and the nature of the relationship between programme commissioners, participants, supposed beneficiaries and evaluators. As discussed in Chapter 4, they have also challenged what is regarded as good evaluation, and the kinds of values and ethical frameworks within which evaluation takes place.

Defining evaluation

Putting a clear boundary around a discrete set of activities which might be described as an evaluation of community-level intervention is also confused by the variety of ways in which the term 'evaluation' can be used. Although many community-level interventions have reports written about them which may be described as evaluations, these would rarely classify as evidence within current thinking, particularly in the health field.

Scriven (1991) in his *Evaluation thesaurus* distinguishes at least four different ways in which the terms 'evaluate' and 'evaluation' are commonly used, the most widely used of which, he suggests:

> 'Refers to the process of determining the merit, worth, or value of something, or the product of that process ... The evaluation process normally involves some identification of relevant standards of merit, worth or value, some investigation of the performance of evaluands on these standards, and some integration or synthesis of the results to achieve an overall evaluation or set of associated evaluations.' (Scriven, 1991: 139)

However, Scriven regrets a growing tendency to use the term 'evaluation' to describe only those activities that are undertaken by a new evaluation discipline that has emerged. This, he suggests, overlooks the many ongoing, self-evaluative activities that are built into most projects and programmes. This kind of self-evaluation activity is often subject to criticism as being unsystematic and unscientific, although such activities often include some level of systematic data collection and analysis.

The complaint about a lack of 'proper evaluation' can also arise from a confusion between evaluation and research. Research is generally seen to be more concerned with generalisability and replicability than evaluation (Scriven, 1991). How far those undertaking the evaluation are seeking to produce generalisable findings relates to the purpose for which the evaluation was undertaken, as well as the methods used. Chelimsky (1997) attempted to cluster the purposes for which evaluation is undertaken under a series of headings, including:

- Accountability
- Development (eg providing evaluative help to strengthen institutions or programmes)
- Knowledge (eg obtaining a deeper understanding of some specific area or policy).

These different purposes are particularly relevant when looking to the evaluations of community-based initiatives to provide the evidence base for further work in this area. As the following section shows, most evaluations of community initiatives in recent years have been undertaken primarily for accountability and development, rather than as a means of creating generalisable knowledge. Although these are valid purposes in themselves, this has led to a style of evaluation, and a lack of publication, which can be very frustrating to those seeking an evidence base for this kind of work which can be set alongside the evidence base of other kinds of health intervention.

Another factor which makes it difficult for existing evaluations to be used in compiling an evidence base for community-level interventions in a conventional health setting is the wide range of research traditions that have helped shape current evaluation practice.

> 'Evaluation in Basil Bernstein's terms (Bernstein, 1971) is a field rather than a discipline. It contains several distinct traditions that at best co-exist uneasily, and at worst vie with or live alongside each other in varying degrees of mutual ignorance. Even these distinct traditions, such as educational evaluation, programme evaluation, technology assessment, quality assurance and management science are disputed, reflecting the full range of contemporary research, philosophies and methodologies: "hard" and "soft", theoretical and atheoretical, positivist and post-positivist.' (Stern, 1993: 76)

As the following section demonstrates, evaluations in the community arena are frequently influenced by emancipatory and participatory research traditions which have a different epistemological base from the experimental research traditions currently favoured in the health field. These different approaches to evaluation reflect other changes that are taking place in the assumptions underlying research practice, as indeed they are changing in the wider scientific community. These include changes such as the following:

From	To
A broadly linear model of causality, which focuses on the links between inputs and outcomes	A growing appreciation and understanding of the complex, holistic and relational nature of reality, as represented by systems, complexity and chaos theory, and non-linear chains of causality. New interest in mechanisms and context as well as inputs and outcomes
A hierarchical approach to research methods in which positivist or objectivist approaches, such as experimental methods, are favoured over other methods	Greater pluralism in which a range of philosophical stances, including scientific realism, pragmatic, constructivist and critical theory standpoints, are accepted equally (although this is more the case in social than in medical sciences)
Knowledge production seen as a professional and academic task undertaken prior to policy and practice development (as represented in most traditional health research and development programmes)	Knowledge production seen as a collaborative, emergent activity resulting from communication and debate between different stakeholders involved in practice on the ground – as represented in action research and newer evaluation models

These changes represent a particular challenge for those seeking to incorporate evaluations based on somewhat different epistemological assumptions into current thinking about evidence-based practice and systematic reviews of evidence. Kelly *et al.* (1993) suggest that these and earlier conceptual shifts described in this section are part of a larger cultural movement from modernity to postmodern thinking. Suggesting that positive health is a postmodern concept, they argue (p. 161) that this is 'not amenable to conventional scientific investigation, or to conventional (modern) scientific discussion', and call for new approaches to research into their impact and effectiveness.

Difficulty in accessing published accounts

Partly the characteristics of community-level interventions, and partly the approaches to evaluation associated with these, made it particularly difficult to access published accounts of evaluations. Initial searches of databases identified several hundred references, but few from the UK. The small number (five) of specific UK examples suggested that few evaluations of projects of this kind have been published in peer-reviewed health journals over the past decade.

This appears to reflect the relatively peripheral place this kind of intervention has occupied until recently in UK health policy, and the low level of resources invested in evaluation. It also reflects the type of evaluations that are typically carried out, and the difficulty in getting such evaluations published. In the era before the advent of journals dedicated to evaluation or health-related audit, it was more difficult to get non-experimental research studies published in peer-reviewed health journals.

Widening the search to include references from bibliographies, an examination of websites and practitioner journals indicated that a great deal of project work had been taking place throughout the 1990s, much of which was subject to some kind of evaluation. Most of the interventions were relatively small scale compared to the large and complex programmes in, for example, the USA or Canada. Although there appears to have been some pressure on projects to undertake evaluations, these were often written up in documents with limited local circulation, and not published for a wider audience. Even the evaluations of relatively large programmes, such as those funded by the Health Education Authority (HEA) (Hills and King, 1993; Cullen *et al.*, 1994), have remained unpublished and difficult to locate. An extensive review of local developments undertaken by the Open University in the early 1990s also remained unpublished (Smithies and Adams, 1993).

As noted in Chapter 2, there has now been a major change in both the kind of work taking place and the kind of evaluation undertaken. However, most of this work is still ongoing and the reports emerging have not yet been published. Some papers are now available through websites (eg HAZnet, www.haznet.org.uk) or conference papers, and these are particularly relevant to the questions of interesting innovations and useful future directions.

Beyond this, there is a considerable body of literature, guidelines to evaluation and debate that could potentially contribute to the picture of appropriate models of evaluation. These are to be found scattered among evaluation journals and websites related to evaluation, community development, regeneration and social exclusion. Through these sources we identified a number of evaluation guidelines that appear

to have had some influence on the practice of evaluation at local level. These include the ABCD handbook (Barr and Hashagen, 2000) and Evaluation of community health development (Luck and Jesson, 1996). However, given the widening circle of domains from which these were obtained, it would be dangerous to claim that this literature review has been anything like comprehensive in identifying all the relevant literature. In many respects, the material presented here should be regarded as illustrative of the different fields, debates and directions relevant to the evaluation of community-based health initiatives, rather than a comprehensive review.

Criteria adopted in selecting literature

Taking these various elements together, the following pointers were used to select the kinds of project or programme, or evaluation debate, to be included in the review.

First, from the articles and reports identified through the search described above (Box 1.1) evaluation reports were selected of health interventions that exhibited some or all of the following characteristics:

- Community focus – involves work either within a limited geographic area, or with a specific interest group, or both (eg might involve work with Asian women within a specific locality)
- Community participation
- Some level of community mobilisation or capacity building
- Supports individual and community empowerment
- Seeks to address health inequalities.

Our notion of what constituted evaluation was broad – we were interested in finding out what projects themselves, and their funders, regarded as evaluations, rather than beginning with a pre-set definition. However, the review did not include economic evaluation; apart from the fact that few completed economic evaluations were located during the period under review, it was felt that this was a specialist area on which the researchers, as non-economists, were not qualified to comment.

The projects and programmes we examined had a broad health focus, even if the health outcomes were secondary to the main aims of the programme (as in New Deal community programmes). However, there was clearly a great deal of literature in the broader field of evaluating community-level interventions that was relevant to the evaluation of specifically health interventions.

This also reflected changes in the policy debate. Community involvement has gained considerable impetus in many policy areas, encouraged in part by the current focus on health inequalities, and in part by a wider democratising agenda. The former has meant that many more programmes addressing issues of poverty and exclusion are now perceived to overlap with those addressing health issues; the latter has led to a much broader interest in public consultation and community involvement in the shaping of all public services – including health services. This has led to developments that span regeneration, service delivery, public health and community development. This is particularly reflected, for example, in the widening scope of primary care to include better community involvement and a larger element of public health work (DH, 1998b, 1999a; Anderson and Florin, 2000).

Overview

This report set outs the findings of the literature review in three main sections. Chapter 2 describes the context in which evaluations in this area have taken place. This is crucial background for understanding the evaluation approach that has been adopted at various points over the past 10–12 years in the UK, particularly noting a change in policy and practice since the change of government in 1997.

Chapter 3 discusses the difficulties encountered in evaluating community-level initiatives, and the various ways in which evaluators before 1997 sought to address these difficulties. This highlights a number of areas in which difficulties remained, which are effectively addressed by existing evaluation designs. Chapter 4 outlines some of the newer evaluation approaches that seek to address these difficulties in new ways. Some of these have been used in the large community-based health programmes established since 1997.

Chapter 5 provides an overview of some of the issues related to quality in evaluation. This includes methodological concerns, as well as the usefulness and appropriateness of evaluation strategies. It also covers issues related to ethics and values. This chapter also discusses mechanisms for quality assurance, including how evaluations are commissioned, and how results have been disseminated and utilised.

Finally, Chapter 6 sets out some of the steps that need to be taken to improve the quality and dissemination of evaluation reports in the future to enhance the evidence base in this area.

2 Contextual factors affecting the evaluation of community health initiatives

Introduction

A good evaluation (as argued in Chapter 4) is one that is useful. However, what constitutes usefulness depends on the kind of evaluation questions being asked. These, in turn, are influenced both by the nature of the community health initiatives themselves, and by the political and professional context within which the questions are being asked.

This section examines some of the developments that have influenced the evolution of community health projects and their evaluation in recent years. A marked change in policy in this area, following the change of government in 1997, is particularly relevant.

Community health interventions pre-1997

In the late 1980s and early 1990s there was a growing interest in addressing public health issues through community-based interventions. Apart from the general interest in community development that had been growing throughout the 1970s and 1980s, two developments were particularly important in encouraging this interest. The first was a series of initiatives, initially begun in Canada and taken up by the WHO's Health for All 2000 and Healthy Cities initiatives, which sought to promote health through a system-wide framework within which multiple interventions could be coordinated at city level (WHO, 1986; Davies and Kelly, 1993; de Leeuw, 2001). These promoted the idea that the community involvement was a key to successful health promotion at the system level. A small number of large-scale, community-wide demonstration projects, particularly in Canada, the USA and Finland, received wide dissemination through their published evaluations. These included the Stamford Three and Five city projects (Farquhar *et al.*, 1985); the Minnesota Health Program (Carlaw *et al.*, 1984); the Pawtucket Heart Health Program (Elder *et al.*, 1986); and the North Karelia Project (MacAlister *et al.*, 1982).

These initiatives stimulated a growing interest in community development in the health field in the UK, leading to the emergence of a new professional group of 'community health workers', and a new division for community development in the HEA. In 1987 the UK Health for All Network was set up, inspired in part by the WHO Healthy Cities movement.

Although a few larger programmes were developed – eg Look after Your Heart community projects (Hills and King, 1993); Heart Beat Wales (Nutbeam *et al.*, 1993); Healthy Sheffield (Thomas, 1996) – the majority of programmes in the UK at this time were relatively small-scale and local. They tended to attract little attention from mainstream health services, and were generally low on the policy agenda. When a white paper, *The Health of the Nation* (DH, 1992) was published, setting out a national strategy for health improvement with specific targets, the community was not given as one of the five settings for intervention. The system-wide approach promoted by the WHO Healthy City initiatives is reflected in an emphasis on 'healthy alliances' and 'healthy partnerships', and such partnerships often included voluntary and community organisations. However, community capacity building, or community development, received little mention in documents such as a guide to the evaluation of healthier alliances (Funnell *et al.*, 1995) or a review of smoking health alliance activities during this period (Evans, 2001).

Despite a lack of support at national level, work at the local level remained quite vibrant, although for the most part relatively small scale and largely funded by local agencies. A review by Smithies and Adams (1990) uncovered 36 projects; by 1997 Sheffield University had established the annual *Community Links* journal dedicated to community health initiatives, and provided examples from over 70 local projects (Community Links, 1997).

The wider context and pressure to evaluate

In researching for examples of evaluations, reviewers found a marked absence of published evaluation reports from the late 1980s and early 1990s. However, information from other sources suggested that this did not reflect an absence of evaluation activity. Community-based projects were often seen by the local health economy as unusual, highly innovative and sometimes controversial (Beattie, 1991b), which meant they were under particular pressure to prove their credentials when it came to applications for funding.

Five main factors influenced the kind of evaluation projects undertaken during this period:

- Location
- Size and duration
- Nature of funding
- Local resistance to evaluation
- Availability (or lack) of support for local evaluation.

Location

Many of the projects taking place in the early 1990s appear to have been viewed primarily as health promotion initiatives in a community setting. This, together with the influence of earlier community-based programmes in the USA and elsewhere, tempted some researchers to adopt an experimental evaluation strategy in order to fit in with the prevailing research paradigm used in the health field to produce evidence of effectiveness.

Other projects and programmes had evolved out of earlier community development programmes, and had a stronger commitment to community participation and involvement. These projects were more likely to undertake evaluations either for purposes of accountability (usually a report of activities with some quantitative data relating to participation), or to use a self-evaluation strategy in support of programme development.

Size and duration

Most projects were small scale, with few resources for extensive evaluation. The lack of national policy interest in this kind of work also counted against any major investment in evaluation. An exception was the relatively large evaluation budget for the National Welsh Heart Health Programme, which enabled evaluators to attempt a relatively sophisticated, quasi-experimental research strategy (Nutbeam *et al.*, 1993). Other evaluations took place at more than one level – eg the Look After Your Heart Community Programme (Hills and King, 1993) had both local evaluations of individual projects and a national evaluation, but all were relatively small-scale exercises.

Nature of funding

Despite the difficulties, projects were often under pressure from their funding bodies to undertake some kind of evaluation of their activities, but this was usually short term. In a guide to evaluation produced in 1996 by Community Health UK, the authors comment that 'public accountability allied to the uncertainty inherent in innovative activities ... means that evaluation is now considered necessary and is required by most funding bodies' (Luck and Jesson, 1996: 10). Evaluation for this purpose often led to evaluation reports that were highly local and specific, and had little circulation beyond the immediate environment of the project or programme and its funding body.

Local resistance to evaluation

The pressure to evaluate activities was not always welcomed at local level. There was concern that the requirement for evaluation was putting increased demands on local people who were already under pressure from other sources. Community Health UK reported that 'many organisations resent evaluation since it seems self-evident that they are tackling important needs and problems in the most effective manner' (Luck and Jesson, 1996).

Availability of support

It was partly in response to these concerns that a number of different organisations sought to provide guidance about evaluation for local organisations. In the early 1990s the Home Office funded a project to support self-evaluation in the voluntary and community sector, while the National Community Health Resource published a bibliography on evaluation. The HEA also provided a list of evaluation resources in a guide to community participation in health promotion (Smithies *et al.*, 1990), and by the mid-1990s Community Health UK had published its guide, *Evaluation of community health development* (Luck and Jesson, 1996). Guidelines to evaluation were also being developed by the Northern Ireland Voluntary Activity Unit and the Scottish Community Development Centre (Barr *et al.*, 1996).

All these resources tended to encourage the use of developmental and participatory evaluation models, which were seen as compatible with the ethos of the programmes themselves (empowering, engaging the community). The evaluation strategies also focused on the contribution of evaluation to the programme or project being evaluated,

and its needs (eg for refunding), rather than on developing a knowledge or evidence base for the wider policy or research community.

Knowledge building was, however, taking place – particularly in the community development arena, where there was considerable discussion of appropriate strategies of interventions and models of best practice, rooted in professional and community experience rather than in research evidence or academic involvement. A similar characteristic is noted by Davies and Kelly (1993) in relation to Healthy City developments during this period which 'grew not as a research project, but as a practitioner-led activity consisting of a series of local community experiments' (Davies and Kelly 1993: 4).

Community health initiatives post-1997

These developments received a considerable boost in the late 1990s, particularly following the change of government in 1997 which signalled a marked change in approach towards public health and community-level interventions. Having been largely ignored during the previous 15 years, health inequalities became a central policy concern following the *Independent Inquiry into Inequalities in Health* (DH, 1998a). The renewed interest in finding ways to address health inequalities paved the way for the implementation of a number of new programmes using community-based and participatory intervention approaches. The community dimension was also highlighted in *Saving lives: our healthier nation* (DH, 1999b) which adopted a more system-based approach to addressing health problems, founded on three-way partnership involving individuals, communities and national government.

These wider policy initiatives were translated into action through the announcement of three major programmes: health action zones (HAZs), healthy living centres (HLCs) and, in a similar and related field, the Sure Start programme. All three represented major investment in tackling health inequalities and all, to a greater or lesser extent, incorporated some level of community involvement.

Although the main focus of the HAZ programme was on joint working between statutory and voluntary organisations within a specific geographic area, a significant amount of the work fitted the general description of community-level interventions (Bauld *et al.*, 2000). Sure Start also has partnership as its major focus, but in addition represents a major investment in local, community-oriented health work with its primary aim of promoting the 'participation of all families in the design and working of the programme' (Henderson *et al.*, 2002).

The HLC programme probably represents the largest major investment in community-level working, with active work with the community being one of the defining characteristics of the 350-odd projects funded across England, Scotland, Northern Ireland and Wales. A DH (1997) circular about the programme emphasised its community orientation: 'Healthy Living Centres will provide opportunities to bring communities together to improve health and wellbeing, and for individuals to take responsibility for improving their own health. Active local participation and a sense of ownership by local people will be key characteristics' (Circular (97) 83, para. 9).

In addition to these large-scale programmes, two other programmes were relevant in terms of understanding community capacity building and community development for health. The Social Action Research Project drew on the experience of local community health development programmes in two cities, to explore the contribution of social capital to community health and wellbeing. The King's Fund Strategic Action Programme for Healthy Communities (King's Fund, 2000) sought to address the question of community capacity building by examining the contextual factors that support or inhibit community development work. This programme aims 'to develop strategies that will improve the ability of the statutory sector to support community-based action for health improvement' (www.kingsfund.org.uk/publichealth/project_summary.html).

There was also considerable work taking place at local level, some of it operating within the context of the Healthy Cities programme. The Good Practice Database on the Our Healthier Nation website (www.ohn.gov.uk) in 2001 listed 79 projects and initiatives under 'health issue – community development'. A survey of local projects compiled by Smithies and Hampson (1999) recorded nearly a quarter of local 'health-related' community-based activities as having 'the community' as their target.

Changing expectations of evaluation

A number of other developments in the past six years at national policy level have had considerable implications for the kind of evaluation expected of health-oriented activities at community level. These include the general expansion in the number of area-based initiatives; changes in expectations around accountability for public sector investment; the growing importance of evidence in the context of policy

development and planning; and a renewed interest in community capacity building in several different sectors.

Growth in area-based initiatives

At around the same time as the programmes described above were set up, the government launched a number of other neighbourhood-based initiatives, stimulated by (among other events) the establishment of the Social Exclusion Unit in the Cabinet Office. When the Regional Co-ordination Unit website (www.rcu.gov.uk) was launched in 2001, many of the 30 area-based initiatives listed either had a specific health brief, or targeted social issues closely related to health. Of the non-health specific initiatives, the one of greatest relevance is the New Deal for Communities, which has health issues as one of the five key areas being tackled in areas of multiple deprivation (DETR, 2000b).

Growing interest in community development

In parallel to these developments, there has been renewed interest in the role of community development. In 1998 the Home Office established the Active Community Unit, with a number of different initiatives to encourage local groups in community-based activities. In the same year, the Charity Commission embarked on a consultation exercise with a view to making community development a charitable purpose in its own right (Charity Commission, 1999). The Community Development Foundation, an active participant in several of the policy action teams established by the Social Exclusion Unit, has been involved in work to develop ways of measuring community capacity building which could be built into wider evaluation strategies (Chanan, 1999).

Changes in expectations of accountability for public expenditure

Another new element in the wider political context has been the growing emphasis on accountability of public expenditure, as represented in the establishment in England of the Better Government and Best Value initiatives, and public service agreements for all areas of government expenditure. This has encouraged a move away from measuring change in terms of input (numbers of staff deployed or services provided) to outcomes, particularly in terms of changes in quality of life in the communities affected, an interest also stimulated by Local Agenda 21 strategies.

These developments have also encouraged a search for more measurable indicators of change (Audit Commission, 2002). This has also stimulated interest in the collection and synthesis of local statistics, a subject that was explored in depth by the Policy Action Team (PAT) 18 working group (DETR, 2000a). The emphasis on planning and performance management, as represented in the modernising government (Cabinet Office, 1999) and Best Value approaches, has also stimulated a demand for greater public participation. Community organisations are expected to play a greater role than previously in the preparation and delivery of strategic plans.

In England the introduction of public service agreements in the health field has led to the establishment of specific targets related to health improvement, as set out, for example, in *The NHS Plan* (2000) relating to cancer, mental health and coronary heart disease, and later for reducing health inequalities (DH, 2001). The focus on targets represented a particular challenge for community-level interventions, which are increasingly viewed in terms of their contribution to the achievement of specific health targets related to morbidity and mortality. In reality, the long timescale required to achieve such specific change, together with their community and positive health focus, means results are rarely measurable by such indicators.

The move to evidence-based policy

A related development has been the growing demand for evidence-based policy. Evidence-based practice in the health field (evidence-based medicine) has been strongly encouraged for some years by the Cochrane Partnerships, the York Centre for Reviews and Dissemination and, most recently, the National Institute for Clinical Excellence. In 2001 the HDA launched its own database to support the building of an evidence base for public health. This new development has spread to other areas, with the Campbell Collaboration for evidence-based social interventions, and the establishment in 2000 of the Economic and Social Research Council (ESRC)'s Centre for Evidence-based Policy at the University of London. The challenge that this evidence-based movement has presented to community-based initiatives lies in the difficulty of applying experimental methods to this type of intervention, as discussed in the following section.

Changes taking place in the wider evaluation community

The evaluation community is itself undergoing a process of evolution, which has also had an influence on wider debates about appropriate evaluation practice. Before the 1990s, evaluation took place in a number of separate 'silos' relating to different domains of interest. In 1996 a new

professional body was set up – the UK Evaluation Society – which reflected the growing self-consciousness of those in this field and a desire to distinguish themselves from those undertaking other kinds of research in similar domain areas. The establishment of a professional association has provided the opportunity for greater exchange between different evaluation traditions, and the dissemination of new models of evaluation from outside the UK. This range of new thinking has dovetailed neatly with the evaluation demands represented by the large-scale, complex programmes set up post-1997. New multi-institute teams have been established to meet the challenge of evaluating these new programmes.

Implications of new programmes for evaluation

The development of new community-oriented programmes at national level has had important implications for the evaluation of activities of this kind.

First, the size of these programmes has meant that they represent a major commitment of public funds, and are seen to warrant a similarly serious investment in terms of evaluation.

Second, unlike previous community-based activities, these programmes are seen as addressing major policy concerns such as social exclusion and health inequalities. Evaluations of these programmes are therefore seen to have much greater public and political visibility, and are under a whole new set of pressures relating to public accountability.

The third aspect concerns the type and structure of the evaluations taking place, which will involve multiple layers and be part of complex, pluralistic research strategies. The main programmes (HAZs, HLCs and Sure Start) are all taking place at several levels and their evaluations reflect this. In addition to national evaluation teams, local projects are also being encouraged to undertake their own evaluations. Some of these are being commissioned externally and involve local institutes of higher education or consultants; others are undertaking self-evaluation, sometimes commissioning specific pieces of work such as community surveys. This multi-layered approach has required mechanisms for linking national and local evaluations closely together. In the case of Sure Start (Henderson *et al*., 2002) and HLCs (Meyrick and Sinkler, 1999), guides to local evaluation were issued at an early stage. These provide an overview of different evaluation strategies that can be used in community-based health projects.

The final aspect relates to the kind of evaluation approach adopted. All the major national evaluations are being conducted by partnerships between two or more different research institutes or academic departments, which represent different skills and experience. These are providing important arenas in which new approaches to the evaluation of community health initiatives can be explored and examined.

Conclusions

In the first part of the 1990s, community-level health interventions were relatively small scale, local activities in which pressure to evaluate came primarily from the need to obtain and retain funding. Although there were a few large national programmes, these were never of the scale found in, for example, the USA and Canada, and there was never a great deal of national policy interest in this kind of work. This meant there was relatively little pressure to use local initiatives as a source of evidence for wider policy developments and little funding invested in evaluation. Research into work of this kind attracted the attention of only a small number of academics with a relatively specialised interest in health promotion or community development.

The emergence of a number of new national programmes following the change of government in 1997 has led to a demand for large-scale evaluations which are being conducted by partnerships between a number of different academic departments and research institutes. This has provided new opportunities for sharing evaluation approaches and disseminating a number of new techniques not previously widely used in this country.

The pressure to evaluate these programmes was also stimulated by a new concern with establishing and measuring specific targets in all policy areas. This encouraged new thinking about outcome measures in the context of community-level interventions, and a fresh examination of the kind of evidence required in areas that were not immediately amenable to experimental research methods. The influence of these various pressures on the kind of evaluation activities undertaken is described in the following chapter.

3 The evaluation challenge

Introduction

Community-level intervention poses a considerable challenge to evaluators. This lies partly in the characteristics of the interventions themselves, and partly in the inadequacy of many of the existing models of evaluation to capture these characteristics at the same time as producing generalisable knowledge that fits with current conceptions of appropriate evidence.

This chapter begins with a brief overview of the characteristics of community-level interventions that create difficulties for evaluators, followed by a summary of approaches used to evaluate these in the past. In outlining the strengths and weaknesses of these approaches, it indicates the areas that more recent evaluation strategies have attempted to address in new ways.

Challenging nature of the interventions

In Chapter 2, an account is given of the characteristics of community health initiatives and how these have been shaped by policy and practice developments. To recap, community-level interventions are seen to be health-related initiatives that:

- Engage the community actively, rather than just being community-based
- Attempt to develop activities from the bottom up, rather than from the top down
- Have a particular value base or rationale for their intervention
- Are oriented to system change as well as individual change
- Involve multiple activities, including infrastructure building
- Incorporate a broad definition of health, often including the concept of positive health rather than disease prevention.

Two other characteristics have emerged, particularly in the current context of multiple, overlapping, area-based initiatives. Interventions are:

- Highly dependent on context
- Rarely the only intervention taking place within an area (particularly in the current policy context).

In reviewing the lessons emerging from various community health projects in the USA, Jackson *et al.* (1989) concluded that programmes of this kind 'tested the limits of all aspects of evaluation research'.

> 'Evaluators face significant challenges in developing sampling, measurement, design and implementation strategies that can survive the complexity of community intervention with enough conceptual and methodological integrity to justify the effort.' (Jackson *et al.*, 1989: 20)

They further argue that, in understanding the impact of an intervention, evaluators need to understand the nature of the community and how the different parts interrelate and influence one another. 'Most important, evaluators should recognise that the introduction of a comprehensive health intervention is a disturbance to the community system since it seeks to change the status quo, reallocate resources and alter existing norms' (Jackson *et al.*, 1989). The relationship between programme interventions, community and evaluators is a dynamic one, subject to change over time and susceptible to influence by many external and internal factors well outside the control of the programme or its evaluation team. The evaluator must 'be prepared for uncontrollable events' (Jackson *et al.*, 1989).

The difficulties faced by community health programmes are shared by other initiatives that seek to incorporate community development approaches. Based on a review of some of the comprehensive community initiatives set up in

the USA to tackle poverty, Connell *et al.* (1995) identify a similar set of difficulties experienced by evaluators, including:

- Resistance to the outcome of the evaluation out of fear that researchers will not do justice to the complex, nuanced, long-term intervention
- Local administrators not cooperating with evaluation efforts
- True experimental designs politically, if not technically, impossible
- Results being used out of context and for political purposes
- Demonstrable programme impacts small compared to programme costs
- Evaluators unable to substantiate with hard figures their intuitive conviction that the programme is having an impact
- Seeking to address health inequalities.

Engaging the community/developing activities from the bottom up

A key element of the challenge to evaluators is the active involvement of community members. This is often seen as a crucial element in projects or programmes aimed at sections of the community that have not been touched by earlier health education interventions, or who have difficulty accessing regular health services. It has also been argued that it is important to the sustainability of the activity once the initial funding has come to an end.

This means that the project or programme is in a constant state of change, responding to the changing characteristics of its participants, and in response to local opportunities as they arise. Change may also arise as national policy priorities change, and funders of the programme and the evaluation may also change their minds about its direction (eg Hills and King, 1993; Bauld *et al.*, 2000). Attempts to standardise the intervention run contrary to the underlying raison d'etre, which relies on flexibility in response to local interest and involvement.

Such programmes often also bring together a range of local players, often in an uneasy alliance between statutory, voluntary and community organisations. Unlike the relatively controlled situation of a clinical trial, different stakeholders may hold different views about the objectives of the programme, as well as about the kind of evaluation strategy that is appropriate and feasible. Who the programme belongs to, and who is authorised to make decisions about the evaluation strategy, may be a source of considerable debate.

Being value-based

Often it is the way such interventions are undertaken, rather than their actual structure, that is a key characteristic. Many community-level interventions start from an empowerment or community development stance which focuses the attention of power politics operating within the programme, and the nature of relationships established between programme providers and recipients. One consequence is that programme providers are often keen to ensure that the value base of the evaluation is similar to that of the programme itself – empowering, and incorporating the views of the community (Barr and Hashagen, 2000).

Kelly *et al.* (1993), reviewing the experience of evaluating healthy cities, make the point that it is the political and value base of community-level interventions that makes them so different from other kinds of health-related intervention. The doctrine of positive health, they suggest, is 'an aesthetic and moral principle, not a scientific one' (Kelly *et al.*, 1993: 165) which can make it difficult to evaluate through conventional scientific methods.

System rather than individual change/involving multiple activities

As noted above, community-level interventions are rarely, if ever, single activities, but are made up of multiple and mutually reinforcing activities occurring simultaneously. For example, a typical project or programme will involve the establishment of some kind of organisational structure, usually involving a small staff group, supported by a committee or partnership of one or more local agencies. In addition to the specific health-related projects or activities (exercise class, stop smoking club, etc) there will also be activities designed to involve the community (possibly as volunteers or as members of the committee, plus via some publicity activities or survey of community views). The larger the programme, the more complex the activities, often with several layers of organisation. In large programmes such as HAZs there are national, area-wide and local projects, each with their own aims and objectives, structures and activities.

In terms of designing an evaluation strategy, this kind of project or programme is very different from the relatively simple input/output model at the centre of most clinical trials. This can lead to a lack of clarity about the unit of intervention under evaluation. Is it the whole programme, its organisation and structure, its projects, or individual activities within the projects?

There may also be a lack of clarity about the aims of the intervention. Is it to establish an infrastructure enabling people to identify and act on health concerns; or to prove that a specific activity, based in the community, will lead to a specific health-related target? Confusing these two very different aims, or having to relate to different stakeholders who believe in different sets of aims, can be particularly disabling for those designing an evaluation.

The lack of published evaluations of healthy cities is attributed by de Leeuw (2003) directly to the complexity of the networks within them, and the different views held about the focus and type of evaluation that should be carried out:

> 'to name a few, academia, citizens, the Healthy City team, research funders, care professionals and politicians may all have their own ideas of what evaluation should pursue. Each of these stakeholders may have very different expectations of the research process and its outcomes.'
> (de Leeuw 2003: 2)

Incorporating a broad definition of health

Most community-level interventions start with a broad definition of the health problem being addressed. Some, like the pioneering programmes in the USA and Finland, begin with a specific health issue, most commonly a heart health agenda. Most community-level interventions are also targeted at achieving positive health interventions – prevention and positive health, rather than cure of illness.

Measuring change when the agenda is broad, the aim is prevention, and the health issue targeted has multiple causes is particularly challenging – a difficulty shared with most public health and health promotion interventions.

In community-led interventions it is especially hard to specify health targets in advance of the action, and this may lead to a gap between the targets initially set and the nature of the intervention adopted. Community participants become involved only when the issue addressed has immediate local interest, and the focus of a programme may change as local interest develops or wanes.

A further difficulty arises because of the timescale of the project and its evaluation – rarely long enough to establish any meaningful changes in health conditions, which may take years to develop. This means that useful interim outcome measures have to be established.

Being context-dependent

Community-level interventions are very context-dependent. The nature of the environment is often a key to which particular combination of activities take place – these are often developed in consultation with the community and determined by local interests and opportunities. The history of the area is often important. For example, in an area where community organisations are well developed, the focus on health may be relatively easy to put into operation. However, where there is little prior infrastructure of community organisation, considerable investment may be required in developing this before the health agenda can be brought into play.

One among a range of interventions

In most cases the intervention being evaluated is only one of a number of factors leading to change. Local attitudes and responses may be affected by changes at national level (a national health promotion programme, or a particular health scare in the media), or by regional or local developments (restructuring of services, or the building of a new hospital). In the current policy environment, areas of deprivation are often the focus of multiple programmes, often with overlapping boundaries. For example, a healthy living centre may be located in the same area as a New Deal community programme, a Sure Start programme and an educational priority area (Bridge Consortium, 2002). This poses an interesting challenge to the evaluator, who is being asked to identify changes that can be attributed to one particular intervention, when in reality any changes are probably due to a synergistic combination of multiple inputs.

Common evaluation strategies

Community-level interventions have generally been under considerable pressure to undertake some kind of evaluation of their activities. Although the review of UK-based evaluations indicated that few studies were being published, information found in practitioner journals, websites and unpublished work suggests there has been a lively debate as to which are the most appropriate evaluation strategies.

Describing and clustering different evaluation activities that have taken place is far from easy. Various classification models have been put forward. Scott and Weston (1998), for example, distinguish between 'empiricist, radical and qualitative' approaches in their review of different evaluation traditions. Rootman *et al.* (2001) comment that researchers were often stereotyped as adopting one of three positions:

strong positivist, constructivist, and a middle position of 'critical multiplism' (see also Cook and MacDonald, 1985). Other distinctions are often made between evaluations that are development-, process- or outcome/impact-oriented, and whether qualitative or quantitative data collection methods are used.

In the present review a fairly clear distinction emerged between evaluations that fit within a broadly experimental paradigm of research, and those that draw on empowerment, participatory or action research traditions. We describe the latter as 'participatory' evaluation approaches. However, in most cases the evaluations reviewed incorporated a broad mixture of different techniques, including hard, quantitative outcome measures (sometimes including data from comparison areas); some more qualitative process indicators; and attempts to incorporate elements of empowerment, or participatory research methods. We describe the last of these as multi-method evaluations.

Evaluation in the experimental research tradition

Experimental designs are widely seen in the clinical field as the most appropriate research design in terms of providing definitive evidence of causality between an intervention and its impact. However, the extension of experimental research designs to community-level interventions has always been controversial.

Connell *et al.* (1995), in their review of the evaluation of community-level interventions in the USA, suggest that inclusion of community-based programmes in the national anti-poverty programmes of the late 1970s led them to be subjected to rigorous research methods. Their raised political profile and demands for measurable programme goals led to a movement from qualitative evaluation approaches to the wider use of experimental designs, often as a condition of funding. A newly established evaluation profession, seeking to enhance its reputation, led many to be attracted by the objectivity, scientific rigour and policy relevance of experimental designs.

It was during this period that some of the most influential community-based health programmes took place (MacAlister *et al.*, 1982; Farquhar *et al.*, 1985; Elder *et al.*, 1986), several being evaluated using quasi-experimental designs. This encouraged researchers to attempt to apply similar methods to the evaluation of community-level interventions in this country.

In the UK, quasi-experimental approaches appear to have worked relatively well where the intervention was fairly specific and the outcomes identifiable at the outset. One such example is the evaluation of 'Burngreave in Action', a community intervention to promote physical activity in two matched inner city areas (Heffron *et al.*, undated). A larger-scale, and more ambitious, attempt was the evaluation of Heart Beat Wales, a national programme to reduce coronary heart disease using community-based interventions (Nutbeam *et al.*, 1993). Another interesting experimental evaluation in another community-based heart health project in England (Baxter *et al.*, 1997) incorporated economic variables in an attempt to demonstrate the cost effectiveness of a range of community-level interventions which did lead to change in diet and smoking in the intervention area.

However, apart from Heart Beat Wales, there is little evidence to suggest that these programmes had more than minor community involvement, providing mainly top-down interventions such as stop smoking support groups, weight control clinics and blood pressure screening. Heart Beat Wales was a much broader initiative that included more by way of capacity building and community involvement activities, which the evaluation sought to capture through combining qualitative data with a quasi-experimental design. Although in many respects a thorough evaluation, the experimental side of the work ran into difficulties when the control sites (similar areas in England) became influenced by the introduction of a national coronary heart disease programme in England.

Participatory and action research-oriented evaluation strategies

One of the difficulties in using an experimental approach was the implicit conflict between the participatory and empowerment-oriented values of community-based programmes, and the research strategy adopted. This could lead to resentment from programme participants and workers, who felt that that the evaluation strategy was failing to appreciate the complex, negotiated and frequently changing nature of the community-level interventions.

Participatory models of evaluation were more appealing to programme participants and workers than experimental methods, because they contributed to the development of the programme and had a closer match with the principles of empowerment and participation which underlay many of the community-level programmes.

This approach to evaluation has been influenced by a number of different research traditions. These include action research, developed by pioneers such as Kurt Lewin and Jacob Moreno earlier in the 20th century (Waterman *et al.*, 2001). This incorporates research into the action taking place, with participants seen as coproducers of the knowledge, and researchers actively involved in the action. The action research tradition has recently gained a more prominent role in the health service, eg in the use of PDSA (Plan, Do, Study, Act) cycles in the Modernisation Agency's 'collaboratives' to promote innovative practice in the health service. In some models of action research the researchers, or those running a programme, retain more control over the variables than in others, as described in the HDA review of participatory approaches to health promotion and health planning (Rifkin *et al*., 2000).

Illuminative evaluation, another influence in participatory evaluation, was developed in the education field and adapted to the evaluation of developmental health initiatives in the late 1970s (Feuerstein, 1978). As with action research, programme participants are seen as coproducers of knowledge, and there is a strong emphasis on flexibility of design and exploration:

> 'Illuminative evaluations are not pre-ordinate, designed in advance with a set of fixed questions or tests to be made. They are far more exploratory in nature, with the problems to be studied being identified through intensive familiarisation with the issues and character of the programme being studied.' (Parlett and Dearden, 1977).

Another influence is fourth-generation evaluation, developed by Guba and Lincoln (1989), which attempts to ground evaluation in a social constructivist rather than positivist epistemology. The assumption that 'facts' and 'values' can be separated, or that any 'objective' account of social actions can be produced, is challenged as Guba and Lincoln argue that social activities consist of multiple realities relating to the different viewpoints or interpretations of different actors involved. The evaluator's role includes ensuring the views of all stakeholders are identified, and supporting negotiation between these different viewpoints in drawing together a common view of the outcome of the programme evaluated.

Although the challenging nature of a constructivist epistemology has received only limited acceptance within the health and health promotion fields, respecting the views and concerns of all stakeholders in programmes being evaluated is referred to in all the standards and ethical guidelines for evaluation described in Chapter 5.

Other influential research traditions in this area have been participatory and empowerment or emancipatory research. In participatory research, the researcher is seen as working in partnership with those affected by the issue being studied, for the purposes of education and taking action or effecting social change. 'When people form a group with a common purpose, to investigate their situation and make decisions ... [they] are transformed – losing fear, gaining self confidence, self esteem, and direction' (Macaulay *et al.*, 1999). On similar lines, in empowerment evaluation, broadly based on Zimmerman *et al.*'s (1992) empowerment theory, participants in a programme are invited to take an active role in evaluation with the aim of fostering self-determination (Fetterman *et al.*, 1996).

The combining of a participant and empowering approach to the evaluation of community initiatives is well illustrated by the ABCD (Achieving Better Community Development) framework produced by the Scottish Community Development Centre (Barr and Hashagen, 2000).

This guide is based on an approach to community learning, which is seen as critical in achieving positive and tangible change for communities and individuals. Community learning involves a wide range of agencies and disciplines, and is seen as contributing to social inclusion, lifelong learning, active citizenship, health improvement and democratic renewal. The approach taken emphasises partnership working and contributing to wider strategies for community participation and planning. This particular community learning approach is based on the values of self-determination, empowerment and participation.

Evaluation as the key to empowerment planning and good practice

Evaluation is too often seen as a negative process, often associated with cuts in funding or closure of projects, too often done at the last minute, and too often seen as a value-for-money exercise. The ABCD approach argues that evaluation is the key to effective practice, and in community development activities it should be conducted in accordance with the values and principles of community development itself. This means working with communities to develop a shared view about what community development interventions are there to do, and how they are going to work towards these objectives together with the community.

Evaluation is seen as an integral part of this community learning, in which both providers and service users take part to achieve continuous improvements in effectiveness and efficiency, and feed lessons back into future planning. The guide focuses on four elements of evaluation: inputs, processes, outputs and outcome.

Multi-method approaches to evaluation

Despite the apparent dichotomy between more hard-edged, experimental research approaches and the more involving participatory approaches, the reality is that many evaluations have sought to incorporate some elements of both traditions. Methods include elements of formative or developmental evaluation which have something in common with the action research tradition – evaluators often gather both quantitative and qualitative data, and may attempt to incorporate some kind of control with which the impact of an intervention is compared.

Such approaches are sometimes described as pluralistic evaluation, and have been strongly recommended by those who have provided advice on evaluation to community-based projects. For example, Beattie (1991a: 27) recommends a portfolio approach in which information is collected and analysed through a diverse range of techniques. He suggests that projects establish 'a systematic collection of records, a cumulative open ended file of many different sorts of information, qualitative and quantitative, representing the broad sweep of the work of the project.' This portfolio project data file would provide a 'key resource, to be raided and edited and put together in different ways, in different formats, for different audiences, for different occasions.' This approach is particularly useful for projects with multiple funding sources, which are sometimes required to produce different evaluation reports for different purposes.

A particularly good example of a pluralistic evaluation is that of Heart Beat Wales (Nutbeam *et al.*, 1993), which incorporated a broad range of measures alongside a broadly experimental design. Another good example of pluralistic evaluation is described by Hunt (1987).

Evaluating a community development project: issues of acceptability

In this project (Hunt, 1987) a health authority set up a research initiative to demonstrate the feasibility and value of using health visitor expertise within a community development framework. Projects included an elderly forum, a fruit and vegetable cooperative, and a tranquilliser support group. A number of different evaluation methods were used in an attempt to assess the impact of the project's presence and activities in the community, including:

- Gathering and monitoring of routine hospital statistics
- Baseline and follow-up community surveys carried out by members of the community who were trained by project staff, and who were paid
- Monitoring of membership of groups, impact on participants and changes over time
- Records of contacts with the project and interaction with networks of contacts both within and outside the community
- Quasi-experimental procedures comparing the effects of various activities on participants.

The multi-method evaluation approach was taken in particular to meet the needs of different stakeholders in the project, including those of the funders, the workers, the community members and the researchers themselves.

Hunt draws attention to the link between the evaluation strategy adopted and the potential for publication. She comments on the researcher's need to 'accommodate all these viewpoints, but still present respectable research, the publication of which will not have an adverse effect on his career.'

Incorporating process and outcome, qualitative and quantitative data

Influenced by the participatory and action research traditions, the use of multiple methods of data collection also enable evaluators to respond to the needs of the programme participants for data on which to base their plans for action. Such methods also tend to pay considerable attention to the process of the programme: how activities are set up and run, and the steps taken to involve other local organisations in partnerships, or to involve community members in running the programme.

Information about process and structures tends to come from qualitative rather than quantitative sources (diaries, interviews, documentary review and observation). Given the complexity of this kind of initiative, qualitative data collection has often been the only option open to the evaluator. Rifkin *et al.* (2000) list aspects of situations that qualitative enquiries can capture, which are difficult to study using quantitative methods. This list includes:

- Collecting detailed data on people's experience, rather than numerical reduction of selected aspects
- Personal contact and insight: involving the researcher with the people being studied to give her/him insight, rather

than refraining from any personal involvement with the subjects
- Design flexibility: adapting inquiry methods as understanding deepens or the situation changes, rather than having a fixed instrument to use on all occasions.

However, there has also been a search for more quantitative process indicators. Rifkin *et al.* (1988) sought to develop a quantitative framework for capturing the nature of participation in five different aspects of community health programmes:

- Needs assessment
- Leadership
- Organisation
- Resource mobilisation
- Management.

Marsden and Oakley (1990) also sought to find ways of measuring community participation, which they describe as being particularly difficult given its unpredictable nature. It can rarely be anticipated in advance how this will manifest in a programme, or with what outcome. However, new approaches to the collection of the more subjective dimensions of community projects in systematic ways are discussed in the following chapter.

Strengths and weaknesses of these approaches

Each of the approaches described above has sought to address the challenge of community-level interventions in different ways. Those adopting experimental and quasi-experimental designs were particularly concerned about producing evidence of effectiveness in order to influence a wider debate about the use of such forms of intervention. This has sometimes worked well with specific activities, where the aims are clear-cut, the intervention itself relatively standardised, and the context relatively unimportant.

However, in the majority of cases these conditions do not apply in community-level interventions. This has meant that the attempt ran into various practical difficulties relating to problems finding an appropriate control group, the lack of standardisation in such interventions, and resistance from those involved in the programme.

There is also an argument that the difficulty with using experimental methods lies deeper than this. Scocozza (2000) argues that the methods incorporate a set of assumptions about the nature of reality which is at odds with those of the intervention. These include the assumption of a relatively passive set of participants, a simple cause–effect relationship between intervention and outcome, and a standardisation of input which runs contrary to the nature of the intervention, which is designed to respond flexibly to local demands. This can lead to a major reality gap between the results of the evaluation, which Hunt (1987), in her review of the evaluation strategy used in a particular community health initiative, describes as a gap between science and reality:

> 'In order to conform to traditional scientific paradigms, evaluation research normally must try to control for extraneous variables, usually by having a control group of some sort, to be able to specify important variables in advance, and to pinpoint "outcomes" which can be assessed in relation to objectives. However, none of these requirements can be adhered to in relation to the types of activities described here. It is difficult, often impossible, to obtain control groups since, apart from the logistics involved, it is antithetical to the philosophy of community involvement that people should be "used", especially at no benefit to themselves. Relevant variables can rarely be specified because of the complexity and dynamic nature of the processes involved. The non-directive nature of the work means that "objectives" are very general to begin with and attain specificity only with the articulated needs of members of the community. There is, so to speak, no advance warning. It is, perhaps, however, the issue of "outcome" where most problems arise, since activities develop, evolve, change direction, ebb and flow. The question inevitably arises, "When is an outcome not an outcome?" '
> (Hunt, 1987)

Participatory evaluation strategies evolved to circumvent many of these difficulties. They enabled the evaluation strategy to be adapted flexibly to the changing circumstances of the programme, and to be undertaken in ways that were in keeping with the underlying philosophy of the programme. Such methods were particularly useful as a strategy for developmental evaluations – contributing to programme planning and reflection. However, they were weak in terms of generating hard proof of the effectiveness of one kind of intervention over another. Evaluators in this tradition have been criticised for 'ducking out' of drawing their own conclusions about the success of a programme, or being reluctant to derive testable hypotheses from particular findings (Pawson and Tilley, 1997).

Pluralistic evaluation strategies have a number of advantages over the previous two approaches. Like participatory approaches, they are usually flexibly adapted to programme requirements and can be changed to take programme developments into account. They also help build bridges between the qualitative and quantitative 'camps' relating to appropriate research designs, and provide a range of data that can be used to fulfil a variety of evaluation purposes.

One of the disadvantages of this approach is that it may fail fully to satisfy either camp. That is, it may result in an evaluation that will serve neither the developmental nor the empowerment needs of the programme being evaluated, nor provide a sufficiently systematic and robust conclusion to meet the needs of those seeking to develop an evidence base for community-level interventions. The York Centre For Reviews and Dissemination's review of evidence to support the government's Our Healthier Nation policy concluded that 'despite suggested benefits of community wellness programmes, current evidence is inconclusive. Studies of community action for health promotion have not been methodologically sound' (York CRD, 2000)

These difficulties are compounded where evaluation resources are limited. If the evaluation strategy is too ambitious, multiple sources of data may be collected, but without a clear framework within which material relating to process in a programme is tied to information about outcomes. It is partly in response to the unsystematic nature of these approaches that a number of new frameworks, such as theory-based and realistic evaluation strategies, and programmatic frameworks (described in the following chapter) have been developed. They help to provide a way of identifying links between programme inputs, context and outcomes, at the same time as going beyond any overly simplistic assumptions about input–output linear causality.

Conclusions

Community-level interventions pose a range of difficulties for the evaluator. Their complex, many-levelled, multi-activity structure means that it is often unclear exactly what is being evaluated, and makes any straightforward link between input and output extremely difficult to establish. This has made it particularly difficult for evaluations in this tradition to be incorporated into an evidence base for decisions about policy and practice in other areas.

Because of the small scale of programmes in the early part of the 1990s, and a lack of policy interest, the evaluation approaches used were primarily designed to support the programmes themselves, either in improving their practice, or accounting for or obtaining funding. These incorporated a number of creative approaches to dealing with the complexity and flexibility of these programmes and activities, and tended to favour either pluralistic evaluation approaches – incorporating a mix of research methods, sometimes based on different underlying epistemologies – or a primarily action-oriented or participatory research approach. The latter involved a close relationship between researcher and researched, and encouraged the inclusion of the subjective views of different stakeholders in programme activities.

A few attempts were made to use experimental and quasi-experimental designs to test systematically the effectiveness of community-level interventions in addressing particular health problems. While these sometimes involve creative solutions to the multiple practical difficulties of implementing such a design, such approaches can, because of their failure to address the complexity of the programme design, also suffer from a reality gap between their findings and the perceptions of those on the ground. They also risk forcing programmes and activities into a level of standardisation that cuts across the principle of local responsiveness and ownership which is part of the rationale behind community-based interventions.

Participatory research designs have been helpful in providing evaluation strategies that contribute to programme development and that reflect the ethos of the programmes being evaluated. However, they have been criticised for failing to produce generalisable conclusions about the effectiveness of such programmes.

Some of the difficulties inherent in these two approaches have been overcome, to some extent, by the use of pluralistic methods that incorporate both qualitative and quantitative data collection, and sometimes also quasi-experimental elements. These can produce a range of research outputs which may satisfy some of the needs for both programme development and generalisable data. However, they can also fail to do either, particularly if resources are very limited. One of the findings of the present review is that many of the pluralistic evaluations undertaken have not been published.

These difficulties, among other causes, have led to a more concerted drive during the past few years to find more effective approaches to the evaluation of complex interventions.

4 New approaches to resolving evaluation dilemmas

Introduction

In Chapter 2 it was suggested that advent of large-scale, large-budget and high-profile community interventions in the health field since 1997 led to pressure to develop evaluation strategies that could generate reliable and useful data about their effectiveness. In responding to this challenge, the evaluation community is currently exploring a number of new evaluation approaches. Some have emerged from within the UK, but in many cases there has been a strong influence from other countries, particularly the USA and Australia, where the evaluation of complex community initiatives is more advanced.

These evaluation approaches generally represent an attempt to find evaluation frameworks that adhere broadly to classic scientific principles, but can accommodate the complexities inherent in community-level interventions. These developments can be divided into five groups:

- New evaluation models that incorporate systematic hypothesis testing, as in experimental research, but that incorporate aspects of theory and context which previous experimental designs tended to overlook
- Development of evaluation frameworks that incorporate a range of different approaches within a single coherent model
- Development of models that seek to address the complexity of modern concepts of health, with their notions of multi-causality, by drawing on complexity theory and socio-ecological models of health
- Development of new outcome measures that seek to capture the system-level changes resulting from community-level interventions, which were previously seen to be more process than outcome
- New approaches to systematic reviews that incorporate a range of research methodologies.

Tackling context and theory

One of the main criticisms of experimental research methods, when applied to community-level interventions, was their tendency to overlook or ignore the theoretical basis of the interventions under examination. These tend to be treated as a black box, with the main focus on the inputs and results, rather than the process of the action. In this respect the experimental paradigm is largely unconcerned with understanding the theory of how the intervention works. The other difficulty was that, in the assumptions behind allocation to intervention and control groups, attempts are made to rule out the influence of variations in individual characteristics and context – but these are the very factors which, in the view of many working in a community context, are the key to the particular development and outcome of their work.

It was in response to these limitations that a new range of theory-based evaluation strategies was developed (Chen, 1990; Chen and Rossi, 1992; Connell *et al.*, 1995; Weiss, 1995). These approaches have much in common with one developed in the UK – realistic evaluation (Pawson and Tilley, 1997). These two innovations – theory-based evaluation and realistic evaluation – seek to address the challenges of complex interventions by applying to experimentation principles drawn from different scientific paradigms. Both have been applied to the two major community-based health programmes – HAZs and HLCs. The Social Action Research Project evaluation also incorporates a realistic evaluation approach.

Theory-based evaluation

The application of theory-based evaluation to complex community initiatives has been explored in depth by the Aspen Round Table in the evaluation of comprehensive community initiatives since the late 1980s (Connell *et al.*, 1995). The approach used by this group is often referred to

as theory of change evaluation, because of its emphasis on the mechanisms of action and the theories underlying these.

Like the action research and participatory evaluation strategies described in the previous chapter, theory of change evaluation requires an active engagement or dialogue between evaluators and programme participants. Part of the task of the researcher is to bring to the surface the theoretical assumptions that different actors in a programme are making about the link between an intervention and its outcomes:

> 'Theory-based evaluation is based on the assumption that social programmes are based on explicit or implicit theories of how and why programmes work. The task of evaluation, in this model, is that the evaluator should help participants in a programme to surface these theories and lay them out in detail, identifying all assumptions and sub-assumptions built into the programme. The evaluation then constructs methods for data collection and analysis to track the unfolding of the assumptions. The aim is to examine the extent to which programme theories hold.'
> (Connell *et al.*, 1995: 67).

The evaluation seeks systematically to test hypotheses based on the theory of change that has been identified through discussions with participants. In this respect the method draws most closely on observational scientific methods, and is broadly in the same hypothetico-deductive tradition as experimental research.

This model provides a tool for separating aspects of a programme that are crucial to its operation from those that are more peripheral, and ensuring these are subject to systematic data collection and measurement. Information about intermediate stages of the action (similar to the process measures described in the previous chapter) is seen as particularly important in establishing links between initial intervention and outcome. This helps to ensure that the success or failure of the final outcome can be correctly attributed to the theorised causes, not to a failure in implementation.

Theory-based evaluation is at the core of HAZ evaluation and one of the national evaluations of HLCs. Bauld *et al.* (2000) noted a number of difficulties in using this approach with HAZs, including:

- Bringing theories in a multi-participant evaluation to the surface: people find the task uncongenial, it is difficult to get agreement, and/or there are political risks in making assumptions explicit, particularly where there are strong racial and ethnic sensitivities
- Measurement: not all the steps or elements in the theory are susceptible to measurement
- Testing theories: these might be too general, or the data may be susceptible to alternative interpretation
- Interpretation: how generalisable are the findings?

Realistic evaluation

Realistic evaluation starts from the assumption that the contexts within which programmes operate are crucial to their outcome. The key question in realistic evaluation is not just whether a particular intervention works, but what works, for whom, and in what context. This emphasises the study of the mechanisms of change – defined as the choices and capacities of those involved in a programme that lead to regular patterns of social behaviour – and of the key elements in the context that help trigger these mechanisms. In a community development context, this means collecting information not only on the activities of the project or programme that is set up, but also on the way this interacts with the particular culture and history of the area in which it is established.

The evaluation needs to address how the context enables or disables the functioning of mechanisms, and to identify what particular context mechanism outcome configurations work:

> 'The basic task of human enquiry is to explain interesting, puzzling, socially significant regularities (R). Explanation takes the form of positing some underlying mechanism (M) which generates the regularity and thus consists of propositions about how the interplay between structure and agency has constituted regularity. Within realist investigation there is also investigation of how the workings of such mechanisms are contingent and conditional, and thus only fired in particular local, historical or institutional contexts (C).'
> (Pawson and Tilley, 1997: 71)

In requiring the evaluator to seek evidence of the mechanisms by which change is brought about, realistic evaluation shares a great deal with the theory of change approach described above. In both the evaluator has a central role to play in teasing out the theory behind the mechanisms used in an intervention:

> 'Realistic evaluators examine outcome patterns in a "theory testing" role. Outcomes are not inspected simply in order to see if programmes work, but are analysed to discover if the conjectured mechanisms/ context theories are confirmed.'
> (Pawson and Tilley, 1997: 215)

As with theory of change evaluation, this means engaging in dialogue with participants of the programme to see what assumptions they have about the mechanisms through which change is to be brought about, and how these are affected by the circumstances in which the programme is set up:

> 'It is not the evaluator's role to provide a theory, but to dig out the programme theories that are often dormant and half articulated – it is the evaluator's task to bring these vibrantly to life.'
>
> 'In order to construct and test context–mechanism–outcome pattern explanations, evaluators need to engage in a teacher–learner relationship with programme policy makers, practitioners, and participants.'
> (Pawson and Tilley, 1997: 203)

This is very much in line with the dynamic relationship between evaluator and project and programme participants that was envisaged by the illuminative, fourth-generation and action research evaluation strategies described in the previous chapter. However, the epistemological basis for these, and for theory of change evaluation, shares a great deal with the experimental approach, particularly in terms of the systematic construction and testing of hypotheses.

Frameworks for programme evaluation

Another advance that has taken place over the past decade is the development of programmatic evaluations, defined as a number of different activities held within a broader framework of intervention or review. This is similar to pluralistic evaluation, as described in the previous section, but with the different elements placed within a more systematic framework which helps identify which evaluation approaches might be used at each stage in the evolution of a programme.

Three frameworks – each drawing on a different research paradigm – were identified as being particularly related to community-level initiatives. These generally relate to the use of different evaluation strategies at different phases of the programme. However, a final consideration is the use of different evaluation strategies at different levels of the programme – taking into account the nature of the new, large-scale programmes operating at both national and local levels.

An experimental design evaluation framework

This framework favours the use of experimental methods in the evaluation of health interventions, but seeks to address the difficulties of using this in a situation in which the interventions are complex and multi-faceted (Campbell *et al.*, 2000). Campbell *et al.* suggest that in such cases it is useful to treat the development of the intervention as a series of distinct stages, similar to the sequential phases of drug development, with different research tasks and various research methods at each stage. They define five steps in the development and implementation of a clinical trial, which they suggest are also applicable to the development of evaluation strategies for more complex health interventions. Different kinds of research activity are seen to be required at each stage in the evolution of what will eventually become an experimental evaluation design. This approach has been built into the recently issued Medical Research Council guidelines for research into complex health interventions, discussed on page 29.

An action learning, policy-related evaluation model

Costings and Springett (1997) developed an evaluation framework for healthy cities, based on action research and drawing on the notion of a policy-making cycle, which the authors define as being 'policy formation, policy implementation and policy outcome'. They see the process of evaluation as a continuous activity consisting of eight steps, each parallel to different steps in the policy-making cycle. They outline a series of evaluation activities which need to be undertaken at each stage of the policy cycle:

- Observe and undertake groundwork
- Establish the purpose of the evaluation
- Design the evaluation protocol
- Decide whether or not to use indicators
- Carry out data collection
- Interpret data obtained
- Provide feedback to people involved
- Undertake actions derived from evaluation findings.

This evaluation approach is designed to be supportive to the development of the programme: flexible and process-oriented. Negotiation between different participants in the programme, using the data derived from the previous stage, is at the heart of the methodology.

A stakeholder evaluation framework based on a programme planning model

Wimbush and Watson (2000) describe an evaluation framework developed and tested by the Health Education Board of Scotland. They link the framework specifically to the growing demand for performance monitoring and more

outcome-oriented evaluation strategies and indicators of effectiveness. The authors identify four main stakeholder clusters: policy makers, programme managers, practitioners and community groups, and users.

The six stages in programme development (based on a number of planning models developed elsewhere) are:

- Planning
- Design and pilot
- Implementation – early start-up
- Implementation – establishment
- Implementation – fully operational
- Dissemination.

Each stage requires different evaluation foci, questions, types and purposes, as well as research questions, applications and questions related to responsibility and the resources needed.

The authors point out that this model demonstrates the need for different forms of evaluation at various stages in the development of effective interventions. They distinguish between project-level self-evaluation (essentially developmental, relating to quality assurance or monitoring) and evaluations that serve a wider function in informing policy and the emerging knowledge base. Making this distinction, they argue, helps to ensure funding bodies are realistic about the kind of evaluation required at different levels, and clarifies the evaluation options available at national and local levels.

Multi-level programme evaluation strategies

This cannot be described as a framework as such, but more as an evolving practice taking place in large-scale programme evaluations. Given the size of programmes such as Sure Start, HAZ and HLCs, there has been a general assumption that evaluation activities will take place at both national and local levels. The extent to which evaluation activities at these different levels complement one another varies from programme to programme. In most programmes the approach to evaluation adopted at local level has tended to be left up to local projects to decide, although often there is some level of support, guidance and training provided from the national team. In the HAZ evaluation, for example, local projects were encouraged to adopt a realistic evaluation framework similar to the national evaluation, with training offered to support this. Similarly, at local programme level support and training was sometimes offered to the very local projects. In practice there was considerable variation in the extent to which this was taken up.

A rather different approach was taken in the HLC evaluation, in which local projects were encouraged to undertake evaluation primarily for developmental purposes, and guidelines were developed in support of this (Meyrick and Sinkler, 1999) before the programme-level evaluation had been commissioned. Difficulties have been reported within other national programmes where national and local evaluations are operating according to very different agendas, and using antithetical approaches (Biott and Cook, 2000). Sanderson (2001) has called for greater understanding of how learning from local and national perspectives can work together, rather than in competition.

Dealing with programme complexity

As noted in the previous chapter, a central difficulty with evaluating community health initiatives is their complexity – multiple interventions taking place at many different levels, often targeted at a range of different outcomes (Bracht, 1990). The increased size of the current programmes, combined with the multiple interventions taking place within the same or overlapping geographical areas, has compounded this difficulty.

In the earlier stages of community health, the complexity of initiatives was tackled through the use of multiple research methods – a pluralistic evaluation approach. Qualitative research methods were employed where a quantitatively based approach would have required such a simplification of the complex variables that the evaluation would have lost its validity. More recent approaches, such as evaluation frameworks, theory-based and realistic evaluation, seek to inject greater rigour into how information about different variables is collected and interrelated. Despite these advances, the evaluation community is still struggling to come to terms with the challenges posed by complex programmes. It is in this context that three other debates are cited here as signalling possible directions for the future.

Medical Research Council guidelines for complex interventions

The Medical Research Council recently sought to provide guidance for research into complex interventions in the health field.

> 'Complex interventions in healthcare, whether therapeutic or preventive, comprise a number of separate elements which seem essential to the proper functioning of the intervention although the "active ingredient" of the intervention that is effective is difficult to specify. If we were to consider

> a randomised controlled trial of a drug vs. placebo as being at the simplest end of the spectrum, then we might see a comparison of a stroke unit to traditional care as being at the most complex end of the spectrum. The greater the difficulty in defining precisely what, exactly, are the "active" ingredients of an intervention and how they relate to each other, the greater the likelihood that you are dealing with a complex intervention.
> (Medical Research Council, 2000: 1)

As has already been noted several times, this difficulty in identifying the key 'active ingredient' is a particular challenge in community interventions, where inputs include structures to develop and support activities, partnerships and community involvement, as well as a range of specific activities.

The Medical Research Council's guidelines suggest a sequence of phases that parallel the steps taken in most drug trials (based on those developed in the Campbell framework above: Campbell *et al*., 2000). These include:

- Pre-clinical or theoretical work – designed to establish the theoretical basis suggesting that the intervention should have the effects expected
- Phase 1 or modelling phase – designed to develop an understanding of the intervention and its possible effects
- Phase 2 or exploratory trial – set up to experiment with the intervention, varying different components to see what effects each has on the intervention as a whole
- Phase 3 – main trial
- Phase 4 or long-term surveillance – to understand the long-term and real-life effectiveness of the intervention.

The guidelines recognise that randomised controlled trials are not possible in all circumstances, but recommend that these steps also be followed in observational trials. They also argue that 'the most challenging part of evaluating a complex intervention – and the most frequent weakness in such trials – is standardising the content and the delivery of the intervention' (Medical Research Council, 2000: 7). This, they suggest, can be overcome through setting acceptable limits within which practitioners can individualise the intervention. However, setting such limits could also be argued to be directly contrary to the bottom-up basis of most participatory community interventions. The guidelines also do little to address difficulties presented by the new understanding of the non-linear and emergent nature of reality provided by complexity theory.

Complexity theory

There is a growing recognition in complexity theory that a key characteristic of complex systems is that they behave in ways that can be highly unpredictable, however much prior information is available. This is because the overall behaviour of a complex system is not simple multiples of the behaviour of individual units, but is 'emergent' as a result of collective behaviour and the relationships between the units. Responses of the whole system to internal or external changes may be minimal, or catastrophic – it is often not possible to predict which.

Complexity theory provides a better basis for understanding the non-linear and often unpredictable processes that take place in most situations in which social policy is implemented, and challenges some of the basic assumptions on which conventional policy evaluation strategies are based (Sanderson, 2001). This is particularly the case where multiple, cross-cutting interventions are taking place simultaneously within the same geographic situation:

> 'Complexity theory has some important implications for evaluation in the context of policy initiatives to address key economic and social problems ... It requires us to recognise that evaluation is necessarily a highly complex endeavour if we accept the realist notion of a multi-layered social reality, the applicability of the concept of dissipative systems and the force of hermeneutic accounts of the role of human agency. It requires us to see evaluation essentially as a craft or "practice" ... comprising a range of methods appropriate to particular circumstances which will provide some understanding of the wider appropriateness of policy initiatives.'
> (Sanderson, 2001: 451)

Although not specifically putting forward an alternative theory of evaluation, Sanderson argues that our growing understanding of complexity challenges some of the assumptions embedded in conventional evaluation practice. He uses this argument to support the need for evaluations to be theory-based and to take account of the wider context in which an initiative takes place, as suggested by theory-based and realistic evaluation strategies. Complexity theory also helps to highlight the importance of a holistic approach to evaluation in which different policies and programmes interact to produce synergies or conflicts. It also emphasises the uncertainty in how policy interventions might or might not work, and the importance of understanding the

processes through which they operate, using qualitative rather than quantitative methods.

Socio-ecological approaches

Complexity theory has also been influential in the development of socio-ecological understandings of health – based on the recognition that health is affected by multiple, interacting factors that have gained increasing support in recent years (including the 1998 World Health Declaration). Initially used in the study of plant and animal populations in their natural habitat, complexity theory was adapted during the mid-1960s and early 1970s to the study of the social, institutional and cultural contexts of people.

While there are several versions of the model (Evans and Stoddart, 1994; Whitehead, 1995; Labonte, 1998), all are based on the notion that the health status of a population is an emergent quality of a whole system of complex interactions. This includes genetic inheritance, the physical circumstances in which people grow up and live (eg housing, air quality, working environment), the social environment (eg levels of friendship, support and trust), and health-related behaviour (eg smoking, alcohol consumption, eating habits, physical exercise). It also includes issues of access to, or lack of, financial and other resources that enable individuals, families and communities to control their lives.

As a way of conceptualising health, and as a model of health intervention, socio-ecological models have considerable advantages over the idea of single causation in health. Because the model encompasses a range of different theoretical and disciplinary perspectives, it can provide a useful basis for cooperation between a range of different disciplines in the delivery of complex health promotion interventions (Stokols, 1996). It is less clear how this model can be used for the evaluation of such interventions, although it is being used as a basis for the selection of indicators for the evaluation of HLCs (Stern, 1999).

New thinking about outcomes and indicators

The previous chapter mentions some of the pressures stimulating debate and discussion about what indicators can be used to measure government targets. There has been a flood of recent documents relating to indicators and output measurement, and it is difficult to do justice to this topic in the context of the present review. For example, in the Audit Commission's consultation on voluntary cross-cutting indicators for local authorities (Audit Commission, 2002), 15 documents are referenced, all produced within the past two years, giving advice and guidance on the development of performance indicators.

However, some important points emerge from these discussions that are particularly relevant to the evaluation of community-based health initiatives. These include the increased availability of performance indicators that local projects might be able to draw on, the growing interest in 'systemic' or structural indicators rather than just measuring individual change, and new thinking about how different kinds of outcome are linked together.

Increased availability of performance indicators

Until very recently it was difficult for a community-level project to gain access to the kind of statistical information that would enable it to assess changes in the local area that might result from its work. However, there is now a great deal of information in the public domain on the kind of performance indicators being used across the country. The Best Value regime for local authorities requires systematic gathering of data for local performance indicators, and this has greatly increased the amount of useful local information. There has been new thinking about how local statistics are collected, such as the work done by the Central and Local Government Information Partnership (CLIP) Task Force on Sustainable Development and the Policy Action Team (PAT) 18 Task Force on Local Statistics (Social Exclusion Unit, 1999). Because many of the important quality-of-life indicators relate to the work of more than one local authority department, there is also growing interest in the development of joined-up indicators drawing information together from a range of different sources (Audit Commission, 2002). The introduction of health improvement areas and performance assessment frameworks in the NHS and social services has led to the development of new performance indicators for health services.

Much of this work is still at an early stage. However, the availability of such information will potentially transform the task of local projects seeking to assess the impact of their work at local level, using population-level or performance indicator statistics.

Growing interest in systemic indicators

Bracht (1990) noted in his review of health promotion interventions at community level that, although these were increasingly being targeted at systems level change (in terms of new structures, enhanced partnership between

agencies, etc), their evaluation still tended to focus on the measurement of impact at individual level. We also noted in the previous chapter that although the process of community-level interventions may provide the key to the outcome, methods for analysing these were often unsystematic, relying primarily on qualitative and descriptive accounts.

Many of the developments taking place in relation to indicators represent a shift from individual to system level. They also represent a move towards a more sophisticated understanding of the multiple factors that affect health, even at individual level. One important contribution to this thinking has come from a growing understanding of social processes and their contribution to health, stimulated in particular by the work of Putnam (1993) on the notion of social capital. This has stimulated the development of survey and research instruments that can capture the extent of an individual's social networks and experience of social support.

There is growing interest in assessing the impact of interventions in terms of an individual's awareness of new community-level structures, their experience of trust, and involvement in their neighbourhood. Measures here include an individual's subjective experience, optimism and sense of belonging to their area, the number of neighbours they recognise and people from whom they can ask small favours (Walker *et al.*, 2000).

At local level there is growing interest in developing tools for measuring community capacity building. This might include surveys of local organisations to assess their number, size and membership, extent of volunteering, and their ability to respond to needs of the area (Chanan *et al.*, 1999). In a later development of this work, Chanan (2002) sets out a framework for measuring community which identifies 16 different variables grouped into six different areas:

- Individual – factors such as self-determination, concern with public issues and level of volunteering or community attitudes
- Community involvement, horizontal – numbers of community and voluntary organisations, social capital and mutual aid
- Community involvement, vertical – voting turnout, response to consultations, extent and effectiveness of community representation
- Services and economic development – extent and range of contribution to public services, social economy and assets
- Inclusion/diversity/cohesion – including measures of extent to which all sections of the population coexist harmoniously and cooperate with one another
- Provision/support/empowerment – community development provision, community and voluntary sector infrastructure and support from partnerships, and public services.

Following the increased awareness of the importance of local infrastructure for the capacities of communities and their groups, there is growing interest in measuring changes in policy, practice and attitudes related to services operating within an area that is the target for change (Chanan, 2002). This links directly to the work outlined above on the development of public service indicators.

One helpful overview of the different kinds of indicators that can be used by community-level health interventions at different levels was provided in the UK by the Evaluation Resource for Healthy Living Centres (Meyrick and Sinkler, 1999). This classifies the kind of indicators that might be used according to three sets of criteria:

- Whether the indicator is a measure of process, impact or outcome
- Whether the indicator measures individual, project- or community-level change
- How the indicator links to the core HLC principles of addressing inequalities, community involvement, partnership, sustainability and supporting local health aims.

Links between different types of outcome indicator

One of the difficulties in earlier evaluation approaches was a tendency to require interventions that were primarily designed to strengthen infrastructure or enhance social interaction, to be measured according to more specific health outcomes.

In an attempt to clarify different levels at which health interventions operate, a recent publication from the International Union for Health Promotion and Education has laid out a framework or outcome model for health promotion activities (IUHPE, 2000). This identifies four levels of outcomes:

- Health and social outcomes
- Intermediate health outcomes (modifiable determinants of health such as healthy lifestyles, effective health services and healthy environments)
- Health promotion outcomes (such as health-related knowledge, attitudes and motivations, social action

and influence, healthy public policy and organisational practice)
- Health promotion actions (education, social mobilisation and advocacy).

In this model, health and social outcomes represent the end point of health and medical interventions, and are usually expressed in terms of mortality, morbidity, disability or dysfunction. However, change may often not be achievable in terms of such an end point, but may be measurable in terms of intermediate outcomes that represent the determinants of these health and social outcomes. Health promotion outcomes are those personal, social and structural factors that can be modified in order to change the determinants of health; they represent the most immediate impact of health promotion activities.

The report argues that the relationship between these different levels of outcome is dynamic rather than linear. The link between them is complex and difficult to trace because the outcomes occur on different timescales and are often separated from actions by a long time period, and because there are often multiple inputs prior to the achievement of a specific outcome. Results may be mediated by a range of factors including the size, comprehensiveness and duration of the intervention, as well as by a wide range of influences not immediately related to the intervention, such as wider changes in culture and social behaviour.

Meta-evaluation and systematic reviews

At the heart of the concept of evidence-based practice is the systematic review. Since the establishment of the Cochrane Collaboration in 1993, there has been a steady refinement of the technology required to compile a systematic review. This has included the development of increasingly refined sets of criteria for the selection of studies, and of quantitative methods for combining findings into one measure of the strength of evidence for the effectiveness of different interventions. These two developments increasingly depend on randomised controlled trials as the gold standard against which all other sources of evidence are assessed, and other evaluation approaches have tended to be excluded. For example, when the York Centre For Reviews and Dissemination undertook a review of research on the effectiveness of health service interventions to reduce variations in health in 1995, it excluded community development-type programmes because:

> 'the quality of evaluations considered (in these and similar areas) was poor, even when the difficulties of evaluating complex interventions given to disadvantaged populations are taken into account.' (York CRD, 1997: 31)

The extension of the idea of evidence-based practice beyond the clinical arena and into such areas as health promotion, public health, education and social care has encouraged further debate about the appropriateness of applying the same principles to all types of intervention. There have been criticisms that too narrow an interpretation of the evidence-based approach has tended to focus attention on the quality of the methodology, often overlooking the quality of the intervention itself (Davey Smith *et al.*, 2001). It has also been argued that this approach can lead to policies that are biased towards interventions that are amenable to experimental evaluation – in other words, it can lead to a bias towards single, top-down, targeted interventions with clear-cut and measurable outcomes, rather than the kind of complex approach represented by community-level interventions (Speller *et al.*, 1997; Gowman and Coote, 1999).

This has led to discussion about the role of other research approaches in systematic reviews. The latest guidelines for conducting systematic reviews by the York Centre for Reviews and Dissemination incorporate quality standards for qualitative research studies, and there is also a Cochrane methods group looking at the role of qualitative research methods in systematic reviews. But the use of a wider range of research methods has implications for the technology used for undertaking the review itself, and has led to closer examination of some of the underlying principles and epistemology of the review.

Realistic synthesis

Examining the underlying epistemology of the review has been part of the focus of Pawson's (2002a) work at the new ESRC Centre for Evidence-Based Policy. Currently, Pawson argues, systematic reviews take one of two approaches. The Cochrane-style numerical meta-analysis is a method based on a successionist notion of causality: the idea that interventions themselves have causal powers. He argues that this works relatively well in relation to clinical research, but the method fails in more complex programmes because it requires too much compression of detail. This can lead to the exclusion of the very information (eg about contextual factors) that would be crucial to any policy decisions about the further use of the intervention:

> 'It is quite possible (indeed quite routine) to have a programme that works well for one class of subjects

> but works against another and whose differential effects will be missed in the aggregation of outputs. This points to the importance of the careful targeting of programmes as a key lesson to be grasped in accumulating knowledge about programmes' efficacy. But this is rather unlikely in meta-analysis, which works to the rival philosophy that the best programmes are the ones that work most widely.' (Pawson, 2002b)

The alternative method, described as narrative review, has more in common with a subjectivist epistemology. This approach is designed to capture more of the detail of the studies under review, usually involving a collation and description of the research available, identification of exemplary case studies, or a synthesis of lessons learned from the most successful interventions in terms of Best Practice conclusions. Such methods operate under what Pawson describes as a configuration approach to causality – that it is a fruitful combination of circumstances that leads to the successful outcome.

However, this approach does not provide a useful basis for generalisation, and often leaves much to the interpretation of the reader. Exemplary cases prove impossible to duplicate, and Best Practice guides are often reduced to a level that is too mundane to provide a realistic guide to practice.

The difficulty common to both numerical meta-analysis and narrative review is that they focus on the question of 'what works?', whereas in reality most interventions only work for some people, some of the time. The key questions are, rather, what works for whom, and under what circumstances?

Pawson therefore proposes the idea of a realist synthesis, based on similar principles to realistic evaluation. This is based on what he describes as a generative causality:

> 'What this tries to break is the lazy linguistic habit of basing evaluation on the question of whether "programmes work". In fact, it is not programmes that work but the resources they offer to enable their subjects to make them work. This process of how subjects interpret the intervention stratagem is known as the programme "mechanism" and it is the pivot around which realist evaluation revolves.' (Pawson, 2002a)

A realistic synthesis takes as its central focus the mechanisms within a programme. It identifies the circumstances under which a particular intervention might or might not work, and tests these against the research results available. This has close links with the theory-based evaluation approaches described above:

> 'Realist synthesis assumes that the transmission of lessons occurs through a process of theory-building rather than assembling empirical generalisations. There is an obvious affinity here with the "theory-driven" approaches to evaluation (Bickman, 1987; Chen and Rossi, 1992; Connell *et al.*, 1995; Pawson and Tilley, 1997). Each of these begins with the notion that programmes are conjectures taking the form, "If we apply programme X this unleashes process Y, which will result in Z". The task of evaluation by these lights is to gather evidence to see if the process occurs as planned and, if it should not, then to amend the theory to account for the divergent outcomes.' (Pawson, 2002a: 347)

Given the complexity and diversity of community-based programmes, this approach provides a promising way ahead for the collation of an evidence base for policy and practice in this area. However, such a method does require evaluation studies that gather and report sufficient information on the interventions taking place and their impact, and requires such studies to be accessible by those undertaking the review.

Conclusions

In the latter half of the 1990s, the arrival of a number of large-scale programmes with a much higher political profile, combined with greater cooperation between different parts of the evaluation community, encouraged the development of new evaluation strategies to address the complexities of community interventions. These approaches pay greater attention to the mechanisms and theory underlying interventions, and to how different contexts can influence the final outcome. Like earlier action research and participatory evaluation strategies, they require active engagement between programme actors, but also encourage a new level of rigour and systematisation in data collection and hypothesis testing. Two of these approaches (theory of change and realistic evaluation) have been adopted in three major programmes set up in the late 1990s that incorporate some elements of community intervention for health – HAZs, HLCs and the process evaluation of the Social Action Research Project.

There has also been a new level of discussion and debate about the nature of meta-evaluations and systematic reviews, which potentially opens the way for a wider range of

evaluation approaches to be incorporated into the evolving evidence base for policy and practice for health promotion. This move will be important for community health initiatives in which experimental and quasi-experimental designs pose a real challenge, and where randomised controlled designs are usually inappropriate. The introduction of the idea of frameworks for evaluation and research also favours a broad range of approaches that might be used in different contexts and at various stages in developing a new way of working or a new programme of activity. Finally, new thinking in the area of indicators and outcome measures provides fresh ways of addressing structural as well as individual health changes that are usually expected only in a community-based project or programme.

5 The changing evaluation context – implications for good practice

Introduction

One of the purposes of this review is to identify what constitutes quality in evaluation design. As previous chapters suggest, a wide range of factors have influenced the kind of evaluation strategies adopted in community-based health initiatives in the UK. This reflects both the different purposes for which evaluation has been undertaken, and the different notions of what an evaluation is.

At the heart of this debate is the fact that evaluations can serve different purposes. The majority of the evaluations examined in the course of this review were undertaken primarily for purposes of programme development or programme accountability. The aim of contributing to a wider knowledge base for activities of this kind was often only a secondary consideration. This meant that research methods were chosen that are appropriate to the former aim, but pose real difficulties in terms of the latter aim – of developing an evidence base for policy in this area according to current definitions of what constitutes an appropriate evidence base.

This highlights the fact that it is the purpose or utility of an evaluation that is a crucial factor in determining what constitutes good practice. Yet, in terms of an evidence base, there has been little systematic study of what kind of evidence people use when establishing a community-based programme of work, or of the impact of different approaches to evaluation. With little consistent information about what has been found to be effective evaluation in different circumstances, it is difficult to come to any firm conclusions as to which evaluation strategies are best.

What this chapter attempts instead is to outline some of the key issues emerging from the literature review, related to the question: what is good quality evaluation in community health initiatives? It also makes reference to some recently developed good practice guidelines that are available for qualitative and action research approaches, as these are widely used in evaluations in this area. This will hopefully prepare the ground for others to undertake a more thorough piece of work in developing a set of principles for good practice in this area.

Establishing quality criteria

Traditionally, the quality of evaluations has been judged primarily in terms of the adequacy of the research design and its methodology: do they answer the question addressed, or not? The examination of methodology, for example, is at the heart of the assessment that is made as to whether or not to include a piece of research in a systematic review.

However, in an evaluation (as opposed to a piece of research; see discussion on page 10) the quality of the methodology adopted is only one of a number of criteria that need to be considered. In various standards for evaluation that have been established by associations of evaluators (although none have been generated by practitioners in the UK), a number of different dimensions have been identified. For example, in 2001 the American Evaluation Association (AEA) set out four core standards in relation to programme evaluations:

- Utility
- Feasibility
- Propriety
- Accuracy.

(AEA website:
www.eval.org/EvaluationDocuments/aeaprin6.html)

As this set of standards is one of the most thoroughly debated sets available, and usefully pulls together many of the key issues related to the evaluation of community health initiatives, it provides the structure for the rest of this section.

Evaluation utility

> 'Utility standards are intended to ensure that an evaluation will serve the information needs of intended users ... information collected should be broadly selected to address pertinent questions about the programme and be responsive to the needs and interests of clients and other specified stakeholders.' (www.eval.org/EvaluationDocuments/aeaprin6.html)

Lack of utilisation of previous evaluations

The usefulness, or otherwise, of the evaluation strategies adopted was highlighted in the early days of community initiatives in the USA, when evaluations often failed to produce answers to the questions that were of importance, either to funders of programmes or local practitioners. One of the consequences of this was that policy decisions about the funding of programmes were often influenced more by political concerns than by the evidence derived from the evaluation of earlier programmes (Connell *et al.*, 1995).

A similar situation appears to have arisen in the UK context. Considerable experience has been gained on the ground and evaluations have been undertaken, but because these are unpublished and inaccessible there remains a major perceived gap between the level of evidence, and the amount of policy interest in establishing new programmes using a community-based approach. The HDA Expert Seminar (organised alongside the present review; see Appendix) noted that policy makers often fail to take into account experience of earlier programme evaluations because of the absence of any organisational memory that such programmes have taken place. A striking example is the Community Development Projects funded by the Home Office from 1969 to1977 (a brief account of this programme and list of research reports is available from the Working Class Movement Library, www.wcml.org.uk), in which there was a major investment in evaluation, but which is rarely referred to in recent times.

However, efforts have been made to overcome this difficulty. There was a systematic attempt to gather evidence together in relation to the main aims of the white paper *Our healthier nation* (DH, 1998a), and a systematic collation of evidence prior to the implementation of the Sure Start programme (Henderson *et al.*, 2002). These reviews have tended to focus on research demonstrating a link between specific health promotion activities and health outcomes – there is rather less information about activities related to community capacity building and community development.

A failure by evaluators to produce usable results was the driving force behind Quinn Patton's (1986) *Utilization focused evaluation*, which urged evaluators to put the usefulness of evaluation results at the centre of strategies adopted. This suggests, among other things, that:

- Evaluations should be use-oriented – aimed at the interest and information needs of specific, identifiable people not 'vague passive audiences'
- There are multiple and varied interests – the process of identifying and organising stakeholders to participate in an evaluation process should be done in a way that is sensitive to and respectful of these various interests
- Evaluators committed to enhancing utilisation have a responsibility to train stakeholders in evaluation processes and the uses of information
- Various factors can affect utilisation – community variables, organisational characteristics, the nature of the evaluator, evaluator credibility, political considerations and resource constraints; the evaluator needs to be sensitive to, and aware of, how various factors and conditions affect the potential use of the evaluation.

Multiple utilities

One of the difficulties in making utility a central standard in the assessment of evaluation quality is that this highlights the fact that stakeholders have different agendas – evaluation can be used for multiple purposes.

> 'The profession of evaluation encompasses diverse perceptions about the primary purpose of evaluation. These include, but are not limited to, the following: bettering products, personnel, programmes, organisations, governments, consumers and the public interest; contributing to informed decision making and more enlightened change; precipitating needed change; empowering all stakeholders by collecting data from them and engaging them in the evaluation process; and experiencing the excitement of new insights.' (www.eval.org/EvaluationDocuments/aeaprin6.html)

This difficulty has been referred to throughout the present review. In Chapter 1 (page 10) we refer to Chelimsky's (1997) clustering of the main purposes for which evaluations are undertaken:

- Accountability
- Development (eg providing evaluative help to strengthen institutions or programmes)
- Knowledge (eg obtaining a deeper understanding of some specific area or policy).

This difficulty is compounded by the fact that the perceived purpose of the evaluation may change as the programme develops, or as new policy priorities emerge. A not untypical example is the Look After Your Heart community projects evaluation (Hills and King, 1993), which began as an exploratory one focusing on deriving learning for future initiatives. However, as the programme evolved the funders became more interested in questions of accountability (whether or not the funding had been well spent) and less interested in learning. The projects funded were also increasingly asked to provide evidence that they were addressing national targets related to the reduction of coronary heart disease. This was a particularly unrealistic expectation, given the small size of the projects and the long timescale required to make any major change in heart disease-related behaviours or morbidity.

This criticism reflected a change in emphasis from bottom-up to top-down, and from learning outcome- to target-driven policy making. One participant in the Expert Seminar captured these shifting concerns of central government:

> 'One year you can go to the Treasury and they will say "We are very interested in learning". That would be in a year when public finances are relatively robust. In another year, it will be "We are not interested in learning, we are interested in efficiency and effectiveness".'
> (HDA Expert Seminar, 2001; see Appendix)

Utility and the role of different stakeholders

The question of whose purpose the evaluation will serve reflects the relative power of different stakeholders, and their capacity to determine whether their purposes will be fulfilled. Barr and Hashagen (2000) identify three groups of stakeholders in most evaluations of community development projects, and their motives (Box 5.1).

It is interesting to note the absence from this list of the wider audience – policy makers or academics – requiring evidence of effectiveness for future policy developments, although these could be seen as programme consumers. These could also be described as the 'vague passive audiences' referred to by Quinn Patton (1986) who are particularly difficult for researchers to incorporate into their thinking in a highly utilisation-focused design.

As is evident from Box 5.1, the interests of stakeholders do not always coincide, and there may be considerable tension within a programme between the different evaluation requirements of each group, as noted as particularly difficult in evaluations of Healthy Cities (De Leeuw, 2003). These various purposes can manifest themselves in a commitment to different evaluation strategies. Participatory or action research-related evaluation approaches are better suited to learning among programme practitioners and participants, while the experimental evaluation models produce more generalisable knowledge. These differences are sometimes recognised in the research briefs for larger programmes (eg for the central and local evaluations of Sure Start and HLCs), which encourage a more participatory, process-oriented evaluation strategy at local level, and a more objective and target-oriented one centrally (as discussed in Chapter 4). The strengths and difficulties inherent in this approach, as used in HAZs, is mentioned in the section on evaluation frameworks (page 28).

Feasibility and appropriateness of the evaluation strategy

A feasible evaluation is one that is 'realistic, prudent, diplomatic and frugal' (AEA website: www.eval.org/EvaluationDocuments/aeaprin6.html). In order to be feasible, evaluations need to be realistic about what they can achieve, and have sufficient resources to achieve these expectations. However – and this is particularly crucial to a community-based programme in which local participation holds the key to success or failure – there also needs to be realism about

Box 5.1 Stakeholders and their interest in evaluation

Parties with an interest in evaluation	**Motives for evaluation**
Programme sponsors: local authorities, central government, trusts, etc	Political accountability for funding, decisions for future funding, lessons for future policy and programmes
Programme workers: field staff and managers	Improving competitiveness in funding market, lessons for future development of practice, personal learning about skills and techniques
Programme consumers: activists, other community members, other agencies	Lessons for future action, personal learning for participants, were aspirations fulfilled?

what exactly the programme is trying to achieve, and the capacity of programme participants to absorb and become engaged with evaluation activities.

Realistic expectations and resources

A common example of a lack of realism in community-level health evaluation is the expectation that a relatively small-scale intervention will have a measurable impact on morbidity or mortality from conditions that take many years to develop, and have many different causes. Projects may be asked to demonstrate links between specific activities and health-related outcomes, when the primary focus of their action is to develop an infrastructure through which activities can be delivered to sections of the population that other approaches have failed to reach. The expectation that small-scale interventions will be able to demonstrate impacts on morbidity and mortality that are well beyond their scope and size was subject to considerable criticism in the Expert Seminar:

> 'Local people should not have to demonstrate that their efforts are having an effect on the bigger questions.'
> (HDA Expert Seminar, 2001; see Appendix).

This issue was also addressed in recent guidance from the Department of Health (DH/Neighbourhood Renewal Unit, 2002) which suggested that when a link is already proven, a local intervention should not be required to demonstrate it again:

> 'Sometimes it is not possible to measure a health outcome directly. Death rates from heart disease are a poor measure of the success of a local strategy to encourage people to take exercise. ... there may be many decades of delay between the intervention and the outcome, which would therefore make it difficult to conclude that one had indeed led directly to the other.
>
> However, it is known that lack of exercise is linked to a higher risk of heart disease. It is sufficient therefore at a local level, to know that more people are putting themselves into a lower risk category by taking exercise, and how often. It is this change that is worth measuring as it is directly and immediately attributable to those actions taken locally which have been expressly designed to lead to health improvement.'
> (DH/Neighbourhood Renewal Unit, 2002: 36)

Unrealistic expectations also arise where evaluations are expected to produce results which would be possible only with a much greater investment of money and expertise. The combination of small evaluation budgets and unrealistic evaluation expectations can be particularly unfortunate. Local evaluators may invest major resources in trying to identify population-level changes when only a relatively small proportion of the population have actually taken part in programme activities, overlooking the very real contribution the activities made to lives of those who were involved (Hills and King, 1993).

Realistic views of the nature of community-level interventions

An evaluation will miss the main point of an intervention if it fails to take into account the underlying rationale for the programme or activities taking place. A common difficulty here is a confusion between the evaluation of specific activities that are part of a programme, and the programme as a whole – which might have been designed as a platform for involving the community in activities. This confusion would be equivalent to confusing an evaluation of a new hospital design with the evaluation of the treatments provided within the hospital.

Most importantly, the evaluation of community-based interventions needs to take into account the level and centrality of community participation in the development and functioning of the programme. Community involvement is essentially a value commitment, as much as an evidence-based practice. Dixon and Sindall (1994) argue that 'The value dimensions (substantive rationality) underpinning programmes and interventions may be overlooked by policy makers, programme planners and professional evaluators under pressure to demonstrate cost effectiveness and efficiency'. They argue that successful and appropriate evaluations of community-based programmes must be sensitive to the needs and success criteria of the communities in which they take place. It helps if there is some shared understanding by researchers, programme sponsors and community actors about the nature of the changes being sought.

Realism about the capacity of the community to engage in evaluation

Successful evaluations require the willingness of those on the ground to cooperate, often being involved in gathering data that will inform the evaluation. However, evaluation activities are often not welcomed, either by the workers involved in a project or the community participants. This reservation is

likely to be particularly marked where the research approach is felt to be out of harmony with the ethos of the project, or to be making unrealistic demands and detracting from what they perceive to be the primary task.

> 'There are communities that have been tramped over by endless rounds of community development projects and other academics looking at it, and so on. We should not waste people's time. It is as valuable as ours.'
> (HDA Expert Seminar, 2001; see Appendix)

The more advanced evaluation strategies described here require considerable involvement from the workers and participants in community projects. They are being asked to articulate the rationale for their project, as well as to provide information about the context and the processes involved in their project activities. Securing this kind of engagement requires considerable skill from evaluators, as well as sensitivity to the underlying dynamics and politics of the activities taking place.

Realism in evaluation practice

Given the high standards set out in the good practice guidelines mentioned elsewhere in this chapter, it also worth injecting a note of caution about what can be achieved in different circumstances. There is a case for asking what constitutes not just a 'high-quality' evaluation, but also a 'good enough evaluation'. In this context it should be noted that:

> 'Evaluation data are never completely clear cut and absolute; studies are always flawed in some way, and there are always questions of reliability and validity. Error free instruments do not and cannot exist in the measurement of complex human, social, behavioural and psychological problems. Evaluation does not provide final answers, but it can provide direction. Thus evaluation does not lead to final statements about causal linkages, but can reduce uncertainty about such linkages.'
> (Quinn Patton, 1986: 151)

Propriety in evaluation practice – the centrality of ethics and values

Propriety standards ensure that an evaluation will be conducted 'legally, ethically and with due regard for the welfare of those involved in the evaluation, as well as those affected by the results' (AEA website: www.eval.org/EvaluationDocuments/aeaprin6.html).

In the evaluation of health initiatives, two conflicting ethical issues have to be reconciled. First, there are the issues surrounding the ethics of introducing new interventions if there is no evidence base related to their efficacy or safety. Second, there are the issues surrounding the conduct of the research – the need to protect the interests of patients or participants from possible harm arising from the work. Concerns about the latter can make it difficult to undertake research to establish the former (Edwards *et al.*, 1998).

Minimising risk

The modernisation of the health service, as in other areas of public expenditure, has led to a growing demand that only practice for which there is a good evidence base should be incorporated, to ensure the effective use of resources and to minimise risk to the public. Unfortunately, there is currently a marked lack of evidence relating to the effectiveness of community-level interventions, and even less regarding any risks attached to this approach. Risks attached to public involvement in community-level interventions have not been subject to discussion, although there is some evidence of risks to health attendant on local-level regeneration initiatives (Popay, 2001). The point was made in the Expert Seminar that:

> 'We should remind ourselves of the damage that has been wrought on communities that have been experimented on, in particular by regeneration over the 40–50 year period since the second world war.'
> (HDA Expert Seminar, 2001; see Appendix)

In the absence of evidence, however, there is a considerable body of professional experience established from years of community development practice, which has accumulated an understanding about what constitutes good and bad practice. Unfortunately this is not always utilised by those drawing up, and working in, new programmes (Hills and King, 1993).

Protection of the interests of those involved in research and evaluation in the community field is complex. In clinical practice, all research protocols need to be passed by ethics committees. The situation as far as ethical approval for evaluations of community initiatives has, to date, been unclear, particularly if these are being run by community organisations rather than the NHS. However, given wider concerns about clinical governance, and the growing popularity of community-level interventions, it is likely that this will emerge as an issue before long. Given the complexity of community-level interventions, there are real dangers that

this could lead to an overly restrictive notion of what kind of research is appropriate for the evaluation of such initiatives, and of the level of informed consent required before people become involved in evaluated activities. This is an area that requires further investigation and debate.

Engaged evaluation strategies and evaluator ethics

As noted in Chapter 4, evaluation strategies are increasingly requiring evaluators to engage directly with programme participants. This is required not only to gather data, but also to tease out underlying theories of change or assumptions concerning the mechanisms through which the interventions are believed to bring about change. In these, as in participatory research strategies, the line between researcher and researched can become quite blurred.

Various codes of professional practice have been developed to address issues related to the ethics of the practice of evaluation, some of which specifically address the nature of the relationship between researcher and researched. The AEA's Guiding Principles for Evaluators has five key principles, including integrity/honesty, respect for people and responsibility for general and public welfare. The Australian Evaluation Society (AES) also has a code of ethics, which includes ensuring evaluators are mindful of the public interest, conduct themselves with courtesy, integrity and truthfulness, and respect any confidentiality undertakings entered into (AES website: www.aes.asn.au/ethics.cfm).

Both sets of guidelines stress the importance of recognising the diversity of interests and values that relate to general and public welfare, and of respecting differences among participants. In the Australian guidelines, 'public interest' includes the 'the full range of stakeholders, and in particular, the potential impacts of differences and inequalities in society'. Given the centrality of health inequalities in community-level interventions, the inclusion of the voice of those often excluded from more conventional health promotion interventions is particularly important.

An important issue arises when the results of evaluations may negatively affect the interests of some of the stakeholders in an intervention, particularly when they affect the continuing funding of a programme which is popular with the community it serves. The AEA guidelines address this by suggesting that 'Evaluators should conduct the evaluation and communicate its results in a way that clearly respects the stakeholders' dignity and self-worth. Conflicts of interest should be dealt with openly and honestly, so that they do not compromise the evaluation processes and results' (AEA website: www.eval.org/EvaluationDocuments/aeaprin6.html). The EU Structural Fund Guidelines on mechanisms for ensuring quality in evaluations also address potential conflicts of interest, and advise that controversial questions should be presented in final reports in a fair and balanced way (EC Structural Fund, 1999).

Ethics and values

There is potentially a tension between the values espoused by the intervention taking place on the ground, and those embedded in the evaluation activity.

This issue is central to the Scottish Community Development Centre evaluation guidelines (Barr and Hashagen, 2000), with similar points made elsewhere in the literature (Yeo, 1993; Fawcett *et al.*, 2001). Insofar as community development is essentially an activity designed to engage the participation of the community, the Scottish Community Development Centre argues that it is 'essential that the community is actively involved in the evaluation though it may not be directly responsible for carrying it out.' It believes that the choice of evaluation strategy and outcome indicators has to incorporate the understanding that evaluation is:

- An integral element of community development
- Fundamental to empowerment
- Part of a learning organisation.

Fawcett *et al.*, (2001) argue that community evaluation should:

- Underline and reflect the issues and the context in which it is happening
- Involve people from throughout the local area
- Link to questions of importance to key stakeholders
- Increase community members' ability to understand what is going on and increase self-determination
- Begin early and be ongoing
- Use its results, if positive, to help sustain and promote widespread adoption of the community initiative and/or its components
- Link with technical assistance to provide total support.

The Scottish Community Development Centre's evaluation guidelines acknowledge that, in some cases, there is a clash of interests between the principles of community development and the principles and values espoused by the communities themselves. This can create a tension for those involved in the evaluation. The guidelines stress the importance of negotiation:

> 'Given that the objectives of communities will not necessarily ... reflect the principles of equity, or be realistic, there will inevitably be need for negotiated agreement on the objectives of the activity by all three parties – communities, sponsors and agencies. Such an agreement is crucial not only for evaluation purposes, but more fundamentally for productive working relationships. We take the view that, while the community should be the primary source of the definition of the criteria for the evaluation of performance, this must be compatible with the overall objectives and values of community development.' (Barr and Hashagen, 2000: para. 2.5.4)

Propriety and care of the evaluator

The AEA guidelines on propriety state that evaluation should be conducted 'with due regard for the welfare of those involved in the evaluation, as well as those affected by the results'. Those involved in evaluation activities can be exposed to risky situations, and it is important to secure the interests of researchers as well as community members. This is particularly the case in participatory evaluation strategies where community members are involved in evaluation activities, sometimes in the role of interviewers in local surveys.

The UK Social Research Association (SRA) has identified the potential risks that can arise for researchers in the field. These include:

- Physical threat or abuse
- Psychological trauma as a result of actual or threatened violence, or the nature of what is disclosed during the interaction
- Finding oneself in a compromising situation in which there might be accusations of improper behaviour.

(SRA website: www.the-sra.org.uk/ethicals.htm)

The SRA points out that such risks can be particularly pronounced where there are local tensions such as strong cultural, religious or racial divisions, or where the topics for discussion may provoke strong feelings in respondents and promote angry reactions. Examples of such topics include poverty, unemployment, relationship breakdown, social exclusion, bereavement and ill health: all issues that can be central in heath-related community interventions. The guidelines argue that it is important to prepare the ground by meeting local community leaders to explain the research and gain their endorsement, and to brief researchers carefully about appropriate social and cultural behaviour.

Methodological competence and accuracy in reporting

The implication of the discussion so far is that methods used in an evaluation relate to the aims of the evaluation (utility value), a realistic assessment of the resources available and the feasibility of undertaking that approach, and an appreciation of the value base of the intervention and its stakeholders. However, once a methodology has been chosen, good practice requires that those undertaking the evaluation are competent in this methodology, and both transparent and accurate in reporting exactly what it is that they did.

The AEA standards require evaluators to 'adhere to the highest appropriate technical standards in conducting their work, whether that work is quantitative or qualitative in nature, so as to increase the accuracy and credibility of the evaluative information that they produce' (AEA website: www.eval.org/EvaluationDocuments/aeaprin6.html). The AES makes the point that evaluators should 'remain current, competent and rigorous in their practice of evaluation' and have a commitment to 'continuous improvement'.

What constitutes high standards will clearly vary from one method to another. Although there have been clear standards set for some time in the area of experimental methods (York CRD, 2001), the availability of similar sets of standards for other methods have been lacking until recently. However, a growing interest in including qualitative research in systematic reviews has led to the development of more structured approaches to the assessment of quality in this type of research, such as guidelines produced by the Cabinet Office (Spencer *et al.*, 2003) and those included in the York guidelines on conducting a systematic review (York CRD, 2001). These address issues similar to those raised elsewhere in this review: the appropriateness of the methodology used, the relationship between researcher and researched (BSA Medical Sociology Group, 1996), the sensitivity of methods to changes occurring during the study (Popay *et al.*, 1998), and the extent to which theory has been used to inform the analysis and any claims to generalisability (Mays and Pope, 1996). The Cabinet guidelines (Spencer *et al.*, 2003) articulate the need for qualitative research to be assessed on its own terms, within premises that are central to its purpose, nature and conduct. They also draw attention to the range of philosophical, ideological and methodological premises that can underlie the use of qualitative methods

which can have different implications in terms of 'quality' indicators. Other issues addressed include questions about the adequacy of reporting and of data analysis, and the extent to which there is sufficient original evidence reported to justify the conclusions drawn (York CRD, 2001).

A set of guidelines has also recently been produced for the assessment of quality in action research (Waterman *et al.*, 2001). This also gives 20 questions that could equally well be used in the assessment of quality in more participatory approaches to the evaluation of community-level health interventions. These, like the quality assessment criteria for qualitative research mentioned above, cover many of the areas listed in this section:

- Is there a clear statement of the aims and objectives of each stage of the research?
- Was the action research relevant to practitioners and users?
- Were the phases of the project clearly outlined?
- Were the participants and stakeholders clearly described and justified?
- Was consideration given to the local context while implementing change?
- Was the relationship between researchers and participants adequately considered?
- Was the project managed appropriately?
- Were ethical issues encountered, and how were they dealt with?
- Was the study adequately funded and supported?
- Were the length and timetable of the project realistic?
- Were data collected in a way that addressed the research issue?
- Were steps taken to promote the rigour of the findings?
- Was the data analysis sufficiently rigorous?
- Was the study design flexible and responsive?
- Are there clear statements of the findings and outcomes for each phase of the study?
- Do the researchers link the data presented to their own commentary and interpretation?
- Is the connection with an existing body of knowledge made clear?
- Is the extent to which aims and objectives were achieved at each stage discussed?
- Are the findings transferable?
- Have the authors articulated the criteria on which their own work is to be read/judged?

(Waterman *et al.*, 2001)

Accuracy

> 'Accuracy standards ... are intended to ensure that an evaluation will reveal and convey technically adequate information about the features that determine the worth and merit of the program being evaluated.' (AEA website: www.eval.org/EvaluationDocuments/aeaprin6.html)

This underlines one of the issues raised by the more recent models of evaluation – the importance of context. The AEA argues that 'the context within which the program exists should be examined in enough detail so that its likely influence on the programme can be identified.'

Transparency

Another point highlighted in all the quality guidelines reviewed is that researchers need to be explicit about their methodology, and on what data they based their conclusions. The AEA also makes the point that evaluators need to say what conclusions they were unable to draw, as well as those that they could. This point is especially pertinent to complex community-based interventions where objectives may be fluid and changing, and achievements hard to measure. Transparency is particularly important when it comes to the question of reporting, and what is included in the report, which is addressed in the following section.

Good practice in commissioning and dissemination

This section focuses on the design and conduct of an evaluation. However, when asking the question 'What constitutes a good evaluation?' it is also important to ask what processes are available to ensure quality in evaluation practice is assured. This question requires evaluation to be seen in a wider context, which includes the processes through which evaluation is funded and results disseminated. A process for quality assurance in evaluation has been outlined by the European Union (EC Structural Fund, 1999). This includes setting the terms of reference for an evaluation, the selection and competence of the evaluation team, the interaction between funders/commissioners and the team, and the production of reports. The process identifies a number of points at which the quality of an evaluation might be assessed, and identifies the different parties involved at each stage. For example, a programme steering group might validate the final report and verify its robustness. Co-decision makers in a programme could be involved in gauging the

reliability and utility of its conclusions and recommendations, and an independent expert could peer review the quality of the methodology.

Good practice in commissioning an evaluation

Difficulties at the commissioning stage was a topic raised at the HDA Expert Seminar. The concern centred on how far the questions evaluators were being asked to address were appropriate or answerable. A parody of the commissioning document of a typical community-based programme was put forward by one participant as:

> 'We wish to hear from applicants who will address this important policy problem. We especially welcome applications from those who will work closely with the local community and seek innovative solutions matched particularly to their local context, and who will work with a wide cross-section of partners, etc' ... I can think of at least three programmes that are currently running in which that entire architecture has been put in place and somebody in discussion with the Treasury has agreed that it will be evaluated using randomised controlled trials.'
> (HDA Expert Seminar, 2001; see Appendix)

The difficulty highlighted by this comment was the mismatch between the programmes being set up to explore a particular issue, and the evaluation strategy that was required in order to (fully) answer the questions policy makers wanted answered. There were particular criticisms of programmes that were set up with impossible goals and timeframes, with lead times that were too short to enable the interventions to come up with realistic plans:

> 'Some of the big interventions that health departments have chosen to invest in, and [research managers] are asked to commission evaluations on, are just fundamentally misconceived. Many of the issues are unevaluable in any meaningful sense.'
> (HDA Expert Seminar, 2001; see Appendix)

The difficulty was felt to have been compounded sometimes by researchers who felt obliged to submit proposals that claimed to be able to answer 'impossible questions' and then found in practice they could not realistically do it. This comment reflects Kelly *et al.*'s reflection on the need for a postmodern approach to evaluation of healthy cities, and their suggestion that such an approach would 'teach us not so much the right answers, but how not to ask the wrong questions' (Kelly *et al.*, 1993: 167)

One solution put forward at the seminar was for evaluators to play a more active role in planning interventions designed to answer specific evaluation questions:

> 'Could we try to design some kind of focused intervention virtually and try and win consensus across a range of communities about what a community-level intervention would be that would have some serious prospect over the next 10–15 years of reducing health inequalities? Can we then get some consensus about how you might commission and construct evaluation to go along with such an intervention ... It would be about trying to enhance the community's capacity to design, implement and learn about complex, messy, big interventions in a different kind of way.'
> (HDA Expert Seminar, 2001; see Appendix)

Good practice for dissemination

The other difficulty, highlighted by the present review, concerned the dissemination of evaluations of community-level interventions. This was twofold: first, the limited circulation and lack of publication of evaluation reports; second, the paucity of information in reports about the exact nature of the programme and its context, information which would be required in order to generalise from the evaluation findings.

Lack of publication

The lack of publication of existing evaluation reports appears to be caused by a combination of factors, including a lack of interest in publication at local level, and a lack of places where evaluation reports that do not fit the general health research paradigm can be posted. On the whole, information on local interventions is circulated through practitioner networks such as Community Health UK, the Community Development Foundation and the Health For All Network. However, these often contain little systematic evaluative information.

There are a number of developments that should lead to an improvement in this situation. These include the HDA's public health Evidence Base database (www.hda.nhs.uk/evidence) and HealthProm*is* (http://healthpromis.hda.nhs.uk), and the Our Healthier Nation database on local initiatives (www.phel.gov.uk/knowledge/practice/practice.html). The success of these will depend on the willingness of local evaluators to make reports available. Guidelines and resources might also be required to advise local evaluators on how to write up their findings in a way that makes them accessible to a wider audience.

Paucity of information

The second difficulty identified here was the failure of many evaluation reports to include sufficient information for the reader to make an informed judgement about the intervention being evaluated, or how it might be replicated. For example, published accounts of experimental evaluations were rich in detail about the evaluation strategy and outcomes, but lacked information about such aspects as the level of community participation, the structure of the activity, and the context within which it was delivered. Non-experimental evaluations often took the form of case studies that provided a detailed description of the intervention and its processes, but lacked detail about the evaluation strategy and programme outcomes.

The situation would be improved if there was a clear set of guidelines available to evaluators on how to document final reports, so that readers can make their own judgements about the results. At a minimum, such reports should include:

- Details about the strategy, including why and by whom this approach was selected, and who was involved in the evaluation
- Process of the activity, including partnership arrangements and links with other initiatives
- Theory or rationale behind the intervention
- How far the community participated in the activities described
- Context in which the activity took place
- Intermediate outcomes, including activities generated, number of participants, extent to which target groups were involved
- Outcomes and impacts, including system-wide changes in community capacity, organisational and service changes, and impact on individual health/quality of life
- How and to whom evaluation results were communicated, and what use was, or was likely to be, made of these findings.

It is less clear whether such a set of guidelines could be implemented in the field. However, if these details were available for a reasonable proportion of the community projects currently taking place, and reports were accessible, many of the difficulties surrounding the compilation of an adequate evidence base for work in this area would begin to be addressed.

Gathering evidence of good evaluation practice

This chapter began with the observation that there is little evidence available on the effectiveness of different evaluation strategies. This was a point taken up in the Expert Seminar, when someone suggested that a specific study be set up to explore how evaluations came to be commissioned and to what extent the results contributed to decisions made by those who commissioned them:

> 'Would it be possible to take a small number of practical evaluation examples and track through why decisions were made? [It would be about asking] ... why was this commissioned in this way? Why did the successful people commissioned choose to adopt this strategy? To what extent did their expectations change during the course of time? What has been their experience of engaging in the field and with advisory committees and sponsors about the marginal choices they were going to make?
> (HDA Expert Seminar, 2001; see Appendix)

The only example of such a study that we were able to find was one in which there were a number of local community heart health projects set up as part of a national programme (Hills and King, 1993). The programme researchers looked at the reason why different strategies had been adopted, and how far it had been possible to carry out the planned evaluations in practice. They concluded that the main influence determining the choice of evaluation strategy was the previous research experience of those commissioning or conducting the evaluation. These included management accounting, epidemiological research, exploratory research, market testing, action research and experimental design. The level of resources available for evaluation activities and the nature of the intervention were also relevant: quasi-experimental approaches were used where the interventions were relatively straightforward, pluralistic approaches in the more complex projects. Self-evaluation was used where few resources were set aside for evaluation.

However, planned evaluation strategies often changed once the projects were established. For example, one project started out from a fundamentally experimental approach, but adopted a case study approach when it became apparent that it was impossible to control for participant characteristics, and the size of groups made it impossible to use the kind of measurement instruments initially planned. Projects introduced more consistent quantitative methods

for measurement of outcomes when under pressure to do so from the wider health community. Although it was apparent in several of the projects that evaluation activities did have an influence on the conduct of activities on the ground, no attempt was made to assess the influence of the evaluation results on wider policy and practice.

Given the number of local evaluations taking place in the context of some of the large programmes (HLCs, HAZs, Sure Start) and the existence of national evaluation teams, there is real potential for undertaking a similar study to the one described above, but on a much larger scale.

Conclusions

This chapter makes the argument that we do not yet fully understand what evaluation strategies are most effective or, indeed, feasible in the context of different kinds of intervention. In the absence of such information, it reviews some of the principles of good practice in evaluation available from various published guidelines, and looks at the relevance of these principles to the evaluation of community health interventions.

These highlight the fact that good quality in the evaluation of community health interventions is only partially attributable to the methodology adopted. While evaluations should be conducted according to the best standards required of the particular research model chosen, there are a number of other issues that need to be addressed in relation to good practice.

A key concern is how far the evaluation strategy is successful in generating useful and usable results. What constitutes 'useful' will depend on the requirements of different stakeholders, and whose needs take precedence. While a qualitative, developmental evaluation strategy might satisfy the needs of programme practitioners and participants, this might not best serve the interests of policy makers and the wider research community that is seeking to establish an evidence base for work in this field. Clarifying the distinction between these two purposes would help ensure evaluation strategies are appropriate to the particular purpose for which they are designed.

The evaluation strategy adopted should also meet the demands of feasibility and propriety. It should be appropriate to the level of resources being committed, and its values and ethics should be broadly consistent with the nature of the intervention itself. In the field of community health interventions, this suggests that evaluation strategies should have the capacity to empower community members, and to provide opportunities for them to participate in the choice of evaluation design and in data gathering and analysis. Evaluations also need to be ethical and respectful. That is to say, evaluators must adopt values of courtesy, honesty and openness, and should take into account diversity in views and opinions. Finally, they must consider the appropriateness of commissioning evaluations in this area, and their wider dissemination.

Unfortunately, the foregoing discussion raises as many questions as it answers – questions that will need to be addressed by researchers together with policy makers and practitioners in the field. Good quality may involve a trade-off between different sets of quality standards, with decisions ultimately being made on the basis of values as much as technical approaches. The list in Box 5.2 highlights some of the questions that need to be addressed in order to make these decisions.

Box 5.2 Questions to be addressed by a good evaluation

Utility

- Has it been made explicit whose purposes the evaluation is serving, and whether its design supports the initiative taking place on the ground? Will it generate new knowledge and/or support the empowerment of the community?
- How will expectations of different stakeholders be mediated, and where are the spaces where these differences can be debated?
- Who will ensure the views of the community are taken into account?

Feasibility

- Are the resources available for evaluation commensurate with the expectations of all stakeholders?
- What has been done to ensure funders and practitioners have a realistic view of the communities they are working with, and their capacity to absorb and engage with different kinds of evaluation activities?
- What has been done to ensure that the models of health outcomes being adopted, and the timescale within which changes might be expected, are compatible?
- Has the evaluation taken into account the historical, political and social (local and national) context in which the programme is taking place?

Propriety

- Has an appropriate ethical framework been developed?
- Who prepared this, and how has it been implemented?
- Has it been determined whether the evaluation strategy/team has a similar value base to the programme being evaluated, and does this matter?

Methodology

- What are the fora in which quality criteria for non-experimental evaluation methods should be debated?
- What are the fora in which newer evaluation methods required for the evaluation of complex, multi-level, multi-objective programmes should be shared?

Commissioning evaluations

- How can the gap between 'impossible questions' and the messy reality of complex, community-based interventions be bridged?
- Should evaluators play a larger role in the design of interventions?

Dissemination and utilisation of evaluation results

- How can evaluators be encouraged to report more fully on their evaluation activities?
- How can more evaluation reports be brought into the public sphere?

Evidence of effective evaluations

- What can be done to gather more information about the effectiveness of different evaluation strategies, including their utilisation and influence on policy and practice at local and national level?

6 Improving the evidence base for community health interventions

Introduction

This chapter seeks to draw together various themes that emerge from this review, and their implications for the future of community-level health evaluations.

A good map of the territory

A number of major advances have been made in recent years in finding ways of combining the rigour of quantitative research methods with qualitative and participatory approaches to the evaluation of community-level interventions. Old dichotomies, which have created difficulties and tensions in the field in the past, are being overcome as new approaches – theory-based and realistic evaluation strategies and programmatic frameworks – provide new ways of combining a range of different research methods within one evaluation. New thinking about outcome measures is providing greater rigour in areas which once relied only on qualitative and subjective reporting. Wider knowledge of complexity theory and a socio-ecological understanding of health are contributing to greater appreciation of the need for methods that reflect the complexity of work at a community level.

However, the Expert Seminar held as part of this review (see Appendix), which brought together many leading researchers in this field, revealed that considerable confusion remains both about the range of evaluation methods available, and where these might be most appropriately applied. This suggests that there is still a lack of clear 'maps' which the policy maker, practitioner or evaluator can use to navigate their way through the field. Various models have been put forward (Billings, 2000), but there is a lack of consensus over terminology. This is partly because a wide range of academic disciplines have contributed to this area of work, bringing different conceptual frameworks and terminologies.

The increased level of interest, and the new academic partnerships that have formed for evaluation of the large programmes established since 1997, provide a fresh opportunity for dialogue and consensus to emerge as to the best way forward. As yet, however, progress has been slow, and it might also be useful for a central agency such as the HDA to take an active role in fostering consensus building, particularly on the range of:

- Different types of community-level intervention
- Different types of evaluation
- Evaluation questions that need to be addressed.

Mapping the different types of community-level interventions

Further work is required to identify the main characteristics of different types of community-level interventions, and how these differ from other kinds of health intervention. This needs to include information about the underlying theory of change, and values informing the work, rather than focusing solely on activities. At present, a danger remains that very different kinds of activity will be seen as equivalent to one another. Two exercise programmes, for example, may be regarded as similar because the activities and health aims are similar, although the processes by which they are established and delivered may be quite different, with different consequences in terms of uptake and impact.

The evaluations of two of the larger community-based health programmes – HAZs and HLCs – have begun to develop maps of the characteristics relating to different kinds of activities and projects taking place at this level (Bauld *et al.*, 2000; Bridge Consortium, 2002). A recent survey of local projects undertaken by Smithies and Hampson (1999) also provides an initial categorisation of different kinds of activities. The aim of developing a map of different types of activity would be to have a common language that

will enable practitioners, policy makers and evaluators to communicate about the kinds of project or programme they intend to set up or evaluate. Such a map is also an important prerequisite for any meta-evaluation or systematic review of work in this area in order to ensure disparate activities are not clustered together inappropriately.

Mapping the different evaluation approaches

The absence of any one, clear classification of different types of evaluation strategy was one of several difficulties that confronted the authors of this review. Although numerous sets of definitions are used to distinguish between different evaluation strategies (Chapter 2), these often overlap and use different terms to describe broadly similar evaluation approaches. This review aims to goes some way toward providing a map of the kinds of evaluation strategy that are particularly relevant to community-based interventions.

However, this can only be a starting point for further discussion and debate. Tensions and debates remain that cannot be ignored. Evaluation models are rapidly evolving, and new strategies may well emerge within a year or two which present fresh solutions to the complexities of evaluation in this field.

Asking the right evaluation questions

In Chapter 4 it became apparent that it is no longer appropriate to ask simply 'what works' when it comes to community-based health initiatives. The kind of intervention that works in one situation may not work in another, and the questions that need to be asked concern which interventions work, in what kind of environment, and why?

It is also apparent that greater emphasis needs to be given to the distinction between evaluation questions that relate to specific activities, and those that relate to structures and processes required to enable such activities to take place. The former are usually relatively straightforward in terms of aims and outcomes, while the latter are much more complex, with questions that concern access and appropriateness of the intervention to the population concerned, as much as specific health outcomes. It is also important to distinguish between evaluation questions that relate to the success of a particular programme (are we achieving our aims?) and those that relate to the development of a body of evidence for further policy development (do such activities achieve similar aims in other circumstances?).

In selecting evaluation questions, two points need to be borne in mind. The first is that evaluation questions should be dictated by the needs of the field, not by the kind of evaluation strategies that are available to provide reliable answers. This may highlight the fact that methods are still not available to provide the kind of bullet-proof evidence of effectiveness that some would like to see.

Second, the identification of appropriate evaluation questions needs to ensure that a variety of views are taken into account. While community-level interventions may provide new ways of enabling health services to meet their targets, they have often been conceived with a somewhat broader notion of health, such as that inspired by the Healthy Cities programme. Programme participants may have different notions from programme funders of the appropriate evaluation questions. While programme funders have a central role in determining the shape of evaluations, a failure to take on the questions of participants may mean that evaluation results fail to convince those whose involvement is central to the whole exercise. This suggests that a radical approach is often required, as indicated by Kelly *et al.* (1993) in relation to healthy cities. Their evaluation, they suggest, must:

> 'attempt to describe the lives of men and women as they live and experience them ... It must grasp the complexities of their thoughts and actions as they move through stressor-rich environments and attempt to come to some kind of accommodation with them ... It will be more akin to the flexible goal-directed micro-analysis of advertising and marketing than to the abstract, principled analysis of science.'
> (Kelly *et al.*, 1993: 167)

Building an evidence base in the field

There is a growing demand for measures of the effectiveness of work in this field to be evidence-based. This requires two developments. First, there has to be agreement about what kind of evidence this will be. This will require more discussion about what meta-evaluation or systematic reviews should be conducted. Second, it will be important to ensure there is a more systematic collation of evaluation reports, so that these can be made accessible to those wishing to undertake meta-evaluations or reviews. Accessibility of evaluation reports remains a difficulty, and the lack of information in published (and unpublished) accounts of evaluations remains problematic. Some initial thoughts on what should be included in accounts of evaluations are presented in Chapter 5, but work will be required to encourage local evaluators to produce reports that include sufficient information in an accessible format.

Ensuring high-quality evaluations

Chapter 5 identifies some of the areas that need to be considered in relation to high-quality evaluation in this field. Initial criteria are highlighted, and these could provide a useful starting point for further work to develop a set of standards specifically for the evaluation of community-level interventions. Particularly promising are new approaches to quality assessment in qualitative and action research strategies.

A set of standards is not enough, however, and there is also a need for quality assurance processes. As indicated in Chapter 5, these are likely to involve not only how evaluations are conducted, but also how they are commissioned, the relationship between commissioners, evaluators and programme participants, and how evaluation outputs are assessed and utilised.

Quality evaluations can be guaranteed only if commissioners are sensitive and aware of the issues involved, and ask questions that are appropriate to the nature of the initiative and the communities with which they are engaged. Chapter 5 points out that there have been few attempts to review the kinds of evaluation strategy that have been adopted, and what their effectiveness or impact has been. The opportunity is now available, through the large number of local initiatives stimulated by national programmes, to assess how evaluations are being shaped and their results utilised. On the basis of this, a more complete picture can be built up concerning what constitutes a truly high-quality evaluation of a community-level intervention.

References

Anderson, W. and Florin, D. (2000) *Involving the public – one of many priorities*. London: King's Fund.

Audit Commission (2002) *Quality of life: using quality of life indicators.* London: Audit Commission. www.audit-commission.gov.uk/reports

Barr, A. and Hashagen, S. (2000) *ABCD handbook: a framework for evaluating community development*. London: Community Development Foundation.

Barr, A., Hashagan, S. and Purcell, R. (1996) *Monitoring and evaluation of community development.* Northern Ireland: Voluntary Activity Unit, Northern Ireland DSS/Scottish Community Development Centre. www.cdf.org.uk/html/pubslist.htm#CommDevInPractice (summary at www.scdc.org.uk/abcd_summary.htm).

Bauld, L., Judge, K., Lawson, L., Mackenzie, M., Mackinnon, J. and Truman, J. (2000) *Health action zones in transition, progress in 2000.* Glasgow: HAZnet. www.haznet.org.uk/hazs/evidence/nat_progress00-sum.doc

Baxter, T., Milner, P., Wilson, K., Freeman, J. and Cooper, N. (1997) A cost effective, community based heart health promotion project in England: prospective comparative study. *British Medical Journal* 315: 582-5.

Beattie, A. (1991a) The evaluation of community development initiatives in health promotion: a review of current strategies. In: *Papers from the OU/HEA 1990 Winter School on Community Development and Health.* Milton Keynes: Roots and Branches, 212-35.

Beattie, A. (1991b) Knowledge and control in heath promotion: a test case for social policy and social theory. In: Gabe, J., Calnan, M. and Bury, M. (eds) *Sociology of health services*. London: Routledge.

Bernstein, B. (1971) The classification and framing of educational knowledge. In: Young, M.F.D. (ed.) *Knowledge and control: new directions in the sociology of education*. London: Collier Macmillan.

Bickman, R. (ed.) (1987) *Using programme theory in evaluation.* New Directions for Programme Evaluation 33. San Francisco: Jossey-Bass.

Billings, J.R. (2000) Community development: a critical review of approaches to evaluation. *Journal of Advanced Nursing* 31 (2): 472-80.

Biott, C. and Cook, T. (2000) Local evaluation in a national early years excellence centres pilot programme: integrating performance management and participatory evaluation. *Evaluation* 6 (4): 399-143.

Bracht, N. (ed.) (1990) *Health promotion at the community level*. Newbury Park, CA, USA: Sage.

Bridge Consortium (2002) *The evaluation of the New Opportunity Fund Healthy Living Centre Programme: first annual report.* www.nof.org.uk

BSA Medical Sociology Group (1996) Criteria for the evaluation of qualitative research papers. *Medical Sociology News* 22.

Cabinet Office (1999) *Modernising government*. Cm 4310. London: HMSO.

Campbell, M., Fitzpatrick, R., Haines, A., Kinmouth, A.L., Sandercock, P., Spiegelhalter, D. and Tyrer, P. (2000) Frameworks for design and evaluation of complex interventions to improve health. *British Medical Journal* 321: 694-6.

Carlaw, R., Mittelmark, M., Bracht, B. and Luepker, R. (1984) Organisation for a community cardiovascular health program: experiences from the Minnesota heart health program. *Health Education Quarterly* 11: 243-52.

Chanan, G. (1999) *Local community involvement: a handbook for good practice*. Dublin: European Foundation for Improvement in Living and Working Conditions.

Chanan, G. (2002) *Measures of community. A study for the Active Community Unit and Research, Development and Statistics Unit of the Home Office*. London: Community Development Foundation Publications.

Chanan, G., Gilchrist, A. and West, A. (1999) *SRB6 Involving the Community*. London: Community Development Foundation Publications.

Charity Commission (1999) *Review of the Register: the promotion of community development*. Discussion document. United Kingdom: Charity Commission. www.charity-commission.gov.uk

Chelimsky, E. (1997) Thoughts for a new evaluation society. *Evaluation* 3 (1): 97-118.

Chen, H.T. (1990) *Theory driven evaluations*. Newbury Park, CA, USA: Sage.

Chen, H.T. and Rossi, P. (1992) *Using theory to improve policy and programme evaluation*. Westport, CT, USA: Greenwood Press.

Community Links (1997) *Community Development and Health: Ideas Annual '97*. Sheffield: University of Sheffield, Community Links website. www.shef.ac.uk/uni/projects/oip/cl/cdhannual.html

Connell, J.P., Kubish, A.C., Schorr, L.B. and Weis, C. (1995) *New approaches to evaluating community initiatives: concepts, methods and context*. Washington, DC: Aspen Institute. www.aspenroundtable.org/vol1/index.htm

Cook, T.D. and MacDonald, M. (1985) Positivism critical multiplism. In: Shortland, L. and Mark, M.M. (eds) *Social science and social policy*. Beverly Hills, CA, USA: Sage, 21-62.

Costings, C. and Springett, J. (1997) Towards a framework for the evaluation of health-related policies in cities. *Evaluation* 3 (3): 345-62.

Cullen, J., Child, C. and Abraham, F. (1994) 'Multi-action in the community'. Final report on the evaluation of the multi-action in the community project, conducted for the Health Education Authority, London (unpublished).

Davies, J.K. and Kelly, M.P. (1993) *Healthy cities: research and practice*. London/New York: Routledge.

Davey Smith, G., Ebrahim, S. and Frankel, S. (2001) How policy informs the evidence: evidence based thinking can lead to debased policy making. *British Medical Journal* 322: 184-5.

DETR (2000a) *National strategy for neighbourhood renewal – report of Policy Action Team 18: better information*. London: Department of the Environment, Transport and the Regions/HMSO.

DETR (2000b) *New Deal for Communities Evaluation Development Project.* London: DETR (Tavistock Institute, Leeds Metropolitan University, Leeds University and Sheffield Hallam University).

DH (1992) *The health of the nation: a strategy for health in England.* White paper (Cm 1986). London: HMSO.

DH (1997) *Healthy Living Centres*. DH Circular MISC (97) 83. London: Department of Health.

DH (1998a) *Our healthier nation: a contract for health*. London: HMSO.

DH (1998b) *Developing primary care groups*. HSC 1998/139. London: Department of Health.

DH (1999a) *Patient and public involvement in the new NHS*. London: Department of Health.

DH (1999b) *Saving lives: our healthier nation*. London: Department of Health.

DH (2000) *The NHS Plan – a plan for investment, a plan for reform*. London: Department of Health.

DH (2001) *NHS Plan – technical supplement on target setting for health improvement*. United Kingdom: Department of Health, Statistical Division and Central Health Monitoring Unit.

DH/Neighbourhood Renewal Unit (2002) *Health and neighbourhood renewal: guidance.* London: Department of Health/Neighbourhood Renewal Unit.

Dixon, J. and Sindall, C. (1994) Applying the logics of change to the evaluation of community development in health promotion. *Health Promotion International* 9 (4): 297-309.

EC Structural Fund (1999) *EANS collection: evaluating socio-economic programmes. Volume 1. Evaluation design and management.* Luxembourg: Office for Official Publications of the European Commission.

Edwards, S.J.L., Lilford, R.J., Braunholtz, D.A., Jackson, J., Hewison, J. and Thornton, J. (1998) *Ethical issues in the design and conduct of randomised controlled trials.* Health Technology Assessments 2 (15). London: National Coordinating Centre for Health Technology Assessment/HMSO.

Elder, J.P. *et al.* (1986) Organisational and community approaches to community-wide prevention of heart disease: the first two years of the Pawtucket heart health program. *Preventative Medicine* 15: 107-17.

Evans, N. (2001) *Tackling smoking through partnerships: lessons learned from the national alliance schemes*. London: Health Development Agency.

Evans, R.G. and Stoddart, G.L. (1994) Producing health, consuming care. In: Evans, R.G., Barer, M.P. and Marmot, T.R. (eds) *Why are some people healthy and others not?* New York: Walter de Gruyter, 27-64.

Farquhar, J., Flora, J. and Good, L. (1985) *Integrated comprehensive health promotion programmes*. Menlo Park, CA, USA: Kaiser Family Foundation.

Fawcett, S.B. *et al.* (2001) Our evaluation model: evaluating comprehensive community initiatives. In: Rootman, I., Hyndman, B., McQueen, D. *et al.* (eds) *Evaluating health promotion approaches*. WHO Regional Publications, European Series, No. 92. Denmark: World Health Organization.

Fetterman, D., Kaftarian, S.J. and Wandersman, A. (1996) *Empowerment evaluation: knowledge and tools for self-assessment and accountability*. London: Sage.

Feuerstein, M.T. (1978) The educative approach to evaluation: an appropriate technology for a rural health programme. *International Journal of Health Education* 21(1): 56-64.

Funnell, R., Oldfield, K. and Speller, V. (1995) *Towards healthier alliances: a tool for planning, evaluating and developing healthy alliances*. London: Health Education Authority.

Gowman, T. and Coote, A. (1999) *Evidence and public health, towards a common framework*. London: King's Fund.

Green, L. and Kreuter, M.M. (1991) *Health promotion planning: an educational and environmental approach*. Toronto/London: Mayfield Publishing.

Guba, E. and Lincoln, Y. (1989) *Fourth generation evaluation*. Newbury Park, CA, USA: Sage.

HEA (1999) *Healthy living centres: guidance to local evaluators*. London: Health Education Authority. www.hda.nhs.uk

Heffron, M., Cochrane, T. and Davey, R. (undated) Burngreave in Action. Sheffield: Institute of Sports Medicine and Exercise Science, University of Sheffield, unpublished.

Henderson, J., Wilkins. K. and Barnes, J. (2002) *Getting started with a local Sure Start programme evaluation*. National Evaluation of Sure Start (NESS).

Hills, D. and King, A. (1993) 'Look After Your Heart in the community: an overview of the LAYH Community Projects Scheme.' Report to the Health Education Authority, London (unpublished).

Hunt, S. (1987) Evaluating a community development project, issues of acceptability. *British Journal of Social Work* 17: 661-7.

IUHPE (2000) *The evidence of health promotion: shaping public health in the new Europe. Book 2: Evidence book*. France: International Union for Health Promotion and Education, report to the European Commission, Brussels/Luxembourg.

Jackson, C., Altman, D., Howard-Pitney, B. and Farquhar, J.W. (1989) Evaluating community level health promotion and disease prevention interventions. In: Braverman, M. (ed.) *Evaluating Health Promotion Programs*. New Directions for Program Evaluation No. 45. San Francisco, CA, USA: Josey-Bass.

Kelly, M.P., Davies, J.K. and Charlton, B.G. (1993) Healthy cities: a modern problem or a post-modern solution? In: Davies, J.K. and Kelly, M.P. (eds) *Healthy cities: research and practice*. London/New York: Routledge.

King's Fund (2000) 'Strategic action programme for healthy communities: draft report on the expert hearing.' http://194.66.253.160/ePublicHealth/html/saphc.html

Labonte, R. (1992) Heart health inequalities in Canada: models, theory and planning. *Health Promotion International* 7 (2): 110-28.

Labonte, R. (1998) A community development approach to health promotion: a background paper on practice tensions, strategic models and accountability requirements for health authority work on the broad determinants of health. In: HEA (eds) *Healthy living centres: guidance to local evaluators*. London: Health Education Authority.

de Leeuw, E. (2001) Global and local (glocal) health: the WHO Healthy Cities Programme. *Global Change and Human Health* 2(1): 34-45.

de Leeuw, E. (2003) *Research issues in the Healthy Cities Project: recommendations for practice*. Maastricht, the Netherlands: WHO Collaborating Centre for Research on Healthy Cities, University of Maastricht.

Luck, M. and Jesson, J. (1996) *Evaluation of community health development.* Bath: Community Health UK.

MacAlister, A. *et al.* (1982) Theory and action for health promotion: illustrations from the North Karelia project. *American Journal of Public Health* 72: 42-55.

Macaulay, A., Commanda, L.E., Freeman, W.L., Gobson, N., McGabe, M.L., Robbins, C.M. and Twohig, P.L. (1999) Participatory research maximises community and lay involvement. *British Medical Journal* 319: 774-8.

Marsden, D. and Oakley, P. (1990) *Evaluating social development projects.* Oxford: Oxfam.

Mays, N. and Pope, C. (1996) *Qualitative research in health care*. London: BMJ Publications.

Medical Research Council (2000) *A framework for development and evaluation of RCTs for complex interventions to improve health.* London: Medical Research Council.

Meyrick, J. and Sinkler, P. (1999) *An evaluation resource for healthy living centres.* London: Health Education Authority.

Nutbeam, D., Smith, C., Murphy, S. and Catford, J. (1993) Maintaining evaluation designs in long term community based health promotion programmes: Heart Beat Wales case study. *Journal of Epidemiology and Community Health* 47: 127-33.

Parlett, M.R. and Dearden, G.J. (eds) (1977) *Introduction to illuminative evaluation: studies in higher education*. Cardiff-by-the-sea, CA, USA: Pacific Soundings Press.

Pawson, R. (2002a) Evidence-based policy. II. The promise of 'realist synthesis'. *Evaluation* 8 (3).

Pawson, R. (2002b) Evidence-based policy: in search of a method. *Evaluation* 8 (2).

Pawson, R. and Tilley, N. (1997) *Realistic evaluation.* London: Sage.

Popay, J. (ed.) (2001) 'Regeneration and health: a selected review of research.' Report to the Nuffield Institute of Health, University of Leeds (unpublished).

Popay, J., Rogers, A. and Williams, G. (1998) Rationale and standards in the systematic review of qualitative literature in health service research. *Qualitative Health Research* 8: 341-51.

Potvin, L. and Richard, L. (2001) Evaluating community health promotion programmes. In: Rootman, I., Hyndman, B., McQueen, D.V. *et al.* (eds) *Evaluation of health promotion: principles and perspectives.* European Series No. 92. Denmark: WHO Regional Publications, 213-40.

Putnam, R. (1993) *Making democracy work*. Princeton, NJ, USA: Princeton University Press.

Quinn Patton, M. (1986) *Utilization focused evaluation.* London: Sage.

Rifkin, S.B., Muller, F. and Bichmann, W. (1988) Primary health care: measuring participation. *Social Science and Medicine* 26 (9): 931-58.

Rifkin, S.B., Lewando-Hundt, G. and Draper, A.K. (2000) *Participatory approaches in health promotion and health planning, a literature review*. London: Health Development Agency.

Rootman, I., Hyndman, B. and McQueen, D.V. (eds) (2001) *Evaluation of health promotion: principles and perspectives.* European Series No 92. Denmark: WHO Regional Publications.

Sanderson, I. (2001) Evaluation in complex systems. *Evaluation* 6 (4): 433-54.

Schein, E.H. (1992) *Organisational culture and leadership.* San Francisco: Jossey Bass.

Scocozza, L. (2000) The randomised trial. In: Gannik, D.E. and Launso, L. (eds) *Disease, knowledge and society*. Denmark: Smafundslitteratur.

Scott, D. and Weston, R. (1998) *Evaluating health promotion.* Cheltenham: Stanley Thornes.

Scriven, M. (1991) *Evaluation thesaurus.* Newbury Park, CA, USA: Sage.

Smithies, J., Adams, L., Webster, G. and Beattie, A. (1990) *Community participation in health promotion.* London: Health Education Authority.

Smithies, J. and Adams, L. (1993) Walking a tightrope. In: Davies, J.K. and Kelly, M.P. (eds) *Healthy cities: research and practice.* London/New York: Routledge.

Smithies, J. and Hampson, S. (1999) *Community participation for health: a review of good practice in community participation health projects and initiatives.* Labyrinth report. London: Health Education Authority.

Social Exclusion Unit (1999) *Better information. Policy Action Team Report 18.* London: Cabinet Office.

Speller, V., Learmonth, A. and Harrison, D. (1997) The search for evidence of effective health promotion. *British Medical Journal* 315: 361-3.

Spencer, L., Ritchie, J., Lewis, J. and Dillon, L. (2003) *Quality in qualitative evaluation: a framework for assessing research evidence*. London: Government Chief Researcher's Office, Cabinet Office.

Stern, E. (1993) Ongoing and participative evaluation: purpose, design and role in the evaluation of a large scale R&D programme. *Research Evaluation* 3 (2): 75-82.

Stern, E. (1999) *Proposal for the evaluation of healthy living centres.* London: Evaluation Research and Development Unit, Tavistock Institute.

Stokols, D. (1996) Translating social ecological theory into guidelines for community health promotion. *American Journal of Health Promotion* 10 (4): 282-98.

Thomas, D.N. (1996) *Uses and abuses in community development: essays from the work of the Community Development Foundation*. London: Community Development Foundation Publications.

Tones, K., Tilford, S. and Robinson, T. (1990) *Health education: effectiveness and efficiency*. London: Chapman & Hall.

Walker, P., Lewis, J., Lingayah, S. and Sommer, F. (2000) *Prove it! Measuring the effect of neighbourhood renewal on local people*. www.neweconomics.org/uploadstore/pubs/.doc_208200094213_ProveIt.doc

Waterman, H., Tilles, D., Dickson, R. and de Koning, K. (2001) *Action research: a systematic review and guidance for assessment*. Health Technology Assessments 5 (23). London: National Coordinating Centre for Health Technology Assessment/ HMSO.

Weiss, C.H. (1995) Nothing as practical as good theory: exploring theory-based evaluation for comprehensive community initiatives for children and families. In: Connell, J. *et al.* (eds.) *New approaches to evaluating community initiatives: concepts, methods and contexts*. Washington, DC, USA: Aspen Institute.

Whitehead, M. (1995) Tackling inequalities: a review of policy initiatives. In: Benzeval, M., Judge, K. and Whitehead, M. (eds) *Tackling inequalities in health. An agenda for action*. London: King's Fund, 22-52.

WHO (1986) *The Ottawa Charter: principles for health promotion*. Copenhagen: World Health Organization Regional Office for Europe.

Wimbush, E. and Watson, J. (2000) The evaluation framework for health promotion: theory, quality and effectiveness. *Evaluation* 6 (3): 1-20.

Yeo, T. (1993) Toward an ethic of empowerment for health promotion. *Health Promotion International* 8 (3): 225-35.

York CRD (1997) *Review of effectiveness of interventions aimed at reducing inequalities in health*. York: NHS Centre for Reviews and Dissemination, University of York.

York CRD (2000) *Evidence from systematic reviews of research relevant to implementing the 'wider public health' agenda.* Contributors to the Cochrane Collaboration and the Campbell Collaboration. York: NHS Centre for Reviews and Dissemination, University of York. www.york.ac.uk/inst/crd/wph.htm

York CRD (2001) *Undertaking systematic reviews of research on effectiveness: CRD's guidance for those carrying out or commissioning reviews*. CRD report no. 4 (2nd edn). York: Centre for Reviews and Dissemination, University of York. www.york.ac.uk/inst/crd

Zimmerman, M., Israel, B.A., Schulz, A. and Checkoway, B. (1992) Further explorations in empowerment theory: empirical analysis of psychological empowerment. *American Journal of Community Psychology* 20 (6): 707-27.

Appendix
Expert seminar: evaluating community-level interventions: where are we now?

A seminar held at the Health Development Agency, 24 April 2001

Proceedings were transcribed from shorthand notes by the Association of Official Shorthandwriters.

Participants

Health Development Agency

Antony Morgan (chair, a.m.)
Professor Mike Kelly (chair, p.m.)
Richard Boreham
Dr Julie Bull
Dr Una Canning
Dr Amanda Killoran
Dr Caroline Mulvihill
Dr Bhash Naidoo
Lorraine Taylor

Tavistock Institute

Vicky Blackburn
Dione Hills
Elliot Stern

Evaluation experts

Dr Vicky Cattel: University College London
Gill Cowburn: University of Oxford
Paula Gaunt-Richardson: Glasgow Caledonian University
Professor Ken Judge: University of Glasgow
Dr Judith Monks: University of Keele
Professor Pauline Ong: University of Keele
Dr Ray Pawson: University of Leeds
Professor Jenny Popay: University of Lancaster
Mai Stafford: University College London
Sylvia Tilford: Leeds Metropolitan University
Dr David Woodhead: The King's Fund

Programme

Introduction
Antony Morgan

The Health Development Agency research strategy
Professor Mike Kelly

Policy context of evaluation – an overview
Elliot Stern

Review of models and methods
Dione Hills and Vicky Blackburn

Theories of change
Professor Ken Judge

Open discussion: chaired by *Dr Ray Pawson*

Taking bite-sized chunks out of the evaluation elephant – lets move the beast on
Antony Morgan

Measuring community heath using the concept of social capital
Richard Boreham

What is good evaluation? Issues and questions
Dione Hills

Open discussion

Summing up and close: *Professor Mike Kelly*

Notes