National Health Demonstration Projects

Evaluation Task Group Review
October - December 2003

Final Report

Scottish Executive, Edinburgh 2004

ISBN: 0-7559-2511-4

Scottish Executive
St Andrew's House
Edinburgh
EH1 3DG

Produced for the Scottish Executive by Astron B3804 6/04

Published by the Scottish Executive, October, 2004

Further copies are available from
Blackwell's Bookshop
53 South Bridge
Edinburgh
EH1 1YS

The text pages of this document are produced from 100% elemental chlorine-free, environmentally-preferred material and are 100% recyclable.

Acknowledgements

We wish to record our thanks to all those members of the demonstration project, evaluation and learning network teams to whom we spoke for their time and hospitality, and for the open spirit in which they responded to our enquiries. We have been very impressed with the positive and constructive way in which they have tackled the difficult issues identified in this report.

Peter Craig	Chief Scientist Office, Scottish Executive Health Department
Margaret Kenicer	Tayside Directorate of Public Health
Cathy Magee	Health Improvement Strategy Division, Scottish Executive
Maria Rossi	Public Health Unit, NHS Grampian (Scottish Executive Health Department secondee)
Patrick West (Chair)	MRC Social and Public Health Sciences, University of Glasgow
Erica Wimbush	NHS Health Scotland

National Health Demonstration Projects Evaluation Task Group

National Health Demonstration Projects
Report of the Evaluation Task Group Review

Contents

Foreword

This report presents the findings of a review of the evaluation of phase 1 of three of the four national health demonstration projects (DPs), Healthy Respect (HR), Have a Heart Paisley (HaHP) and Starting Well (SW). It is important to note that in order to inform decision-making our review had to be carried out before any of the external evaluation final reports had been submitted to the Scottish Executive, so has not taken into account the content of those reports. Some evidence on early outcomes in SW was available at the time of writing together with interim reports on the process of implementing SW and HaHP, but the timing meant that it has not been possible to fully address the first objective of the group's remit. Since completing the main text of our report we have had sight of drafts of two of the three final reports. While these confirm our view that the impact of the DPs is currently questionable, and that their potential effectiveness has been limited by a variety of methodological problems, it is also the case that the evaluations have provided much useful evidence for learning. Moreover as a group we feel this sequence of events provides further evidence to support our conclusion that more careful consideration of timing and better integration with decision-making processes is essential to make best use of the resources invested in evaluation.

National Demonstration Projects Evaluation Task Group

Section 1: Introduction

1.1 The National Health Demonstration Projects

In 1999, the public health White Paper *Towards a Healthier Scotland* pledged £15 million over 3 years to support four National Health Demonstration Projects (DPs) in priority areas of child health, sexual health of young people, coronary heart disease and cancer. The purpose of the DPs was to act as a testing ground for national action and to provide a learning resource for the rest of Scotland. They were intended to combine the best existing evidence with innovative practice, a concept that has been a matter of ongoing debate throughout the life of the projects. In 2003, *Improving Health in Scotland – The Challenge* announced the intention to support a second phase of the demonstration project programme, for up to a further 3 years, subject to a review of existing interventions and investments.

This report considers the evaluation of the first three years (Phase 1) of the three health demonstration projects, Healthy Respect (HR), Have a Heart Paisley (HaHP) and Starting Well (SW), with the aim of making recommendations for the design and commissioning of future evaluation work. We do not consider the Cancer Challenge which was commissioned and evaluated separately but a summary of the pilot, evaluation models, methods and results is included at Annex A.

1.2 Purpose of this Report

The National Steering Group for the Health DPs (now the Co-ordinating Group) set up four task groups, one of which was the Evaluation Task Group. Its remit is:

- to make a methodological assessment of the Phase 1 evaluations (outcomes and process) in order to develop an account of the strength of available evidence;
- to make recommendations on the basis of available evidence for further evaluation and policy recommendations;
- to report to the National Health Demonstration Projects Co-ordinating Group.

Members of the group have reviewed published research relating to community-based sexual health, heart health and early years interventions, and documents relating to the process of commissioning both the DPs themselves and the independent evaluations. The group has met the evaluation teams and members of the management teams for each of the three projects, as well as the learning network co-ordinators. This report describes our findings and makes recommendations for the evaluation of Phase 2 of the demonstration projects and for the evaluation of future initiatives of this kind.

Our remit required us 'to develop an account of the strength of available evidence'. That is to say, our assessment depended not just on how well the evaluations were designed and conducted but also on a) the strength of the existing evidence base, and b) how closely the DPs (and therefore the evaluations) matched the evidence base. We have therefore considered the prior evidence for the effectiveness of community based heart disease, sexual health and early years interventions in some detail, rather than restricting our review to materials originating from the demonstration project programme.

1.3 Structure of Report

The remainder of this report is divided into four sections:

- The commissioning process
- The evidence base of the demonstration projects
- Evaluation in practice

- Conclusions and recommendations

1.4 The Commissioning Process

1.4.1 The demonstration projects

As stated above, the Government's intention to set up four national health demonstration projects (DPs) was announced in the 1999 public health White Paper, *Towards a Healthier Scotland*. Bids had already been invited for 'The Cancer Challenge', a screening programme for the early detection of colorectal cancer. Expressions of interest in bidding for the three remaining projects, which were to focus on health promotion and protection in the period leading up to birth and the first five years of life, promotion of responsible sexual behaviour among young people, and the prevention of heart disease, were invited in May 1999. Two hundred and fifty-eight notes of interest were received and six organisations were invited to develop full proposals. These were assessed by a panel chaired by the Chief Medical Officer comprising senior officials from the Scottish Executive Health Department and Health Education Board of Scotland (HEBS), and three independent experts. The three successful bids were announced in February 2000 after the formulation of more substantive proposals and post-tender negotiations – Have a Heart Paisley (HaHP), Healthy Respect (HR) and Starting Well (SW). The DPs are described in detail in Section Two. Bids were not expected to set out detailed plans for evaluation, and this was not a criterion by which they were assessed.

1.4.2 The external evaluations

External evaluations of the three DPs were commissioned through a selective tender process in which six organisations were invited to bid for one or more of the evaluation studies. This 'selective tendering' method is the usual procedure for procuring social research; it is less often used, though far from unknown, in health research. Tenders were invited after the demonstration project contracts had been awarded, and the successful proposals formed part of the documentation sent to prospective bidders. A number of those invited to tender declined on the grounds of lack of capacity, or because they had concerns about the 'evaluability' of the projects. This reluctance to tender may be indicative of the limited research capacity in Scotland, but could also be because evaluating community-based interventions presents many evaluation problems, and does not suit traditional methods or produce tidy publications. The fact that many academic units in Scotland and elsewhere refuse to engage in evaluating such interventions means in turn that there is a limited national evidence base available for such practical interventions.

One bid was received for the evaluation of each demonstration project. Each bid was evaluated by a tender panel comprising representatives of the Chief Scientist Office (CSO), the Central Research Unit and Public Health Policy Division of the Scottish Executive Health Department (SEHD) and HEBS (now NHS Health Scotland)

The invitation to submit notes of interest told demonstration project bidders that they could not have access to research funds to develop their applications, and also that they should concentrate on disseminating good practice because there would be separate, external evaluations. This, coupled with the fact that the evaluations were commissioned separately, after the demonstration project contracts had been awarded, helped to cement a distinction between the implementation and the evaluation of the DPs which has been problematic for both sides. This is discussed further in Section 3.

The objectives of the evaluations were extremely broad. They required the evaluators to develop methods which would provide an overall estimate of the success or failure of each demonstration project:

> The aim of the evaluation is to determine how far the projects meet their own aims and objectives, and to use this information to judge their contribution towards the overall goals of the demonstration project initiative. ...
>
> The overall range of objectives for each project is broad. Some of the proposed health outcomes will respond slowly, even if the interventions are effective. Some may be straightforward to measure, but many raise difficult methodological issues associated with identifying and attributing effects. Each set of aims and objectives contains outcome as well as process elements, and the evaluation must cover both, in order to describe and explain the projects' impact. In addition, each project is meant to improve health within its local target population, and to demonstrate methods that could be adopted elsewhere. Both aims are relevant to the evaluation. (Invitation To Tender, April 2000)

The original brief was very open with regard to research methods:

> Bidders are invited to propose methods for the evaluation. These will vary between demonstration projects, and between objectives, but it is envisaged that a combination of quantitative and qualitative methods will be required, using primary data gathered via surveys and interviews, and secondary analysis of monitoring data. Methods should be rigorous, precise and appropriate to the hypotheses being tested, taking account of the particular problems associated with evaluation of area-based interventions. They should be efficient in terms of the sampling and other requirements placed on the demonstration projects. They should be geared towards estimating the effectiveness of the projects in meeting their objectives, and understanding the reasons for the success or otherwise of the projects. The overall conceptual and methodological framework should be made explicit. Tenders should indicate clearly which outcomes they propose to measure, and give reasons for the choice. If surrogate outcome measures are proposed, evidence of their validity should be offered. (Invitation To Tender, April 2000)

Nevertheless, the fact that the commissioning process was managed by the CSO, on behalf of the Public Health Policy (now Health Improvement Strategy) Division of the SEHD may have given rise to certain expectations about how the evaluations should be conducted. It was suggested to us that the CSO's involvement ultimately encouraged a view that a traditional approach relying on controlled comparisons should be adopted.

Section 2 The Evidence Basis of the Demonstration Projects

Each of the DPs was developed within the context of an evidence base relating both to general issues, notably the effect of deprivation on health, and to the effectiveness of community-based interventions proposed. In this section we consider how strong the evidence – including new evidence emerging during the projects' lifespan - was in relation to each DP and how closely the DPs matched the evidence-base. Any assessment of the effectiveness of DPs should take account of these issues since a failure to demonstrate might reflect the weakness of the initial evidence, rather than a failure of implementation.

2.1 Starting Well (SW)

2.1.1 The demonstration project

SW aims to demonstrate that child and family health can be improved via an intervention comprising two integrated components: a programme of intensive home-visiting combining parental support and parenting education, and a strengthened network of community-based support services for children and parents. The parenting education programme was not specified in the original bid, but following a review of several possibilities, the 'Positive Parenting Programme' (Triple P), developed by Sanders et al. (2000) in Australia, was chosen by SW and has been delivered both on an individual and group basis. The dual approach, combining family and community perspectives, reflected a belief that by adopting a broader ecological model of health, the impact of the home visiting programme is likely to be enhanced.

The project is located in two disadvantaged communities in Glasgow. To date, 1555 families with a child born from December 2001 to December 2003 have participated (98% of the total number of births). 19% of participants were recruited (and home visited) in the ante-natal period, the remainder (81%) having their first home visit soon after the birth of the child. The intervention consisted of regular visits from health visitors supplemented by lay 'health support workers' up to the child's third birthday. One of the areas (Glasgow South) contains a relatively high proportion (18%) of families of South Asian origin, reflecting one of the demonstration project's aims; that is, to address the needs of black and ethnic minority groups.

2.1.2 The evidence base

The original proposal for SW was accompanied by a literature review of the evidence relating to the role of poverty for child and maternal health and implications for parenting together with an assessment of the evidence on the effectiveness of family interventions in ameliorating these effects. Almost all the available evidence on interventions was derived from randomised controlled trials (RCTs), involving allocation of families into an intervention (nurse visiting) and control arm (no nurse visiting). Two systematic reviews of those studies were also referred to. While the weight of evidence was, and remains, supportive of the SW rationale, a number of points, some of which were mentioned in the original SW literature review, are worth noting here.

- Both systematic reviews (MacMillan et al., 1993; Hodnett & Roberts, 1999) commented on a range of methodological problems associated with the evaluation studies, including small sample sizes, inadequate randomisation, lack of blinding in evaluation (Johnson et al., 1993) and problems of attribution (intervention or extraneous factors). Even in the Elmira trial (Olds et al., 1986), methodologically judged the best (MacMillan et al.,

1993), the investigators acknowledged that an intervention effect could be attributable to pre-existing differences between groups and/or reporting bias.

- Both systematic reviews offered only cautious support, each noting variability in results between studies and between outcomes in the same study. Some studies, including one in England (Lealman et al., 1983), demonstrated no intervention effect, or effects not in the expected direction. Many of the reported differences did not reach conventional levels of statistical significance, and some were misinterpreted. For example, the SW literature review reported the results of a meta-analysis as showing 'a significant preventive effect of home visiting on the occurrence of childhood injury' (p.9) while the original review drew attention to the fact that the pooled odds ratio of 6/8 trials reporting lower injury encompassed 1.0 (Hodnett & Roberts, 1999). The error in the SW review was corrected in the later evaluation proposal, the finding being described as a 'non-significant trend' (Gray et al., 2000, p.4).
- Almost all studies referred to were restricted to first born children, and most focussed on more deprived populations. Support for intervention effects is often found in specific sub-groups of study participants, typically the most vulnerable. Thus, in the Elmira trial, at age two, significant differences between nurse visited and control groups were found only among young (<19) unmarried mothers of low socioeconomic status (SES) (Olds et al, 1986). In a long-term follow-up (age 15), an intervention effect on delinquency and substance use was similarly restricted to this group (Olds et al., 1997). In another trial of a 'therapeutic programme' to promote maternal competence and parenting, the effect of the intervention was observed only among low IQ mothers, higher IQ mothers doing better with a traditional programme (Barnard et al., 1988). These results generally underpin recommendations to target interventions either at vulnerable families in the community (Gutelius, 1977) or (less usually) to 'communities with high concentrations of low-income married women' (Olds et al., 1997, p.642). All available evidence, however, refers to RCTs of families rather than communities.
- The best evidence for intervention effects are also associated with particular characteristics of the intervention: namely early (ante-natal) contact, frequent home visiting by nurses in both the ante-natal and post-natal period and long term (up to 5 years) contact (MacMillan et al., 1993). There was also some evidence that programmes with a 'therapeutic' ethos are more effective (Barnard et al., 1988; Gutelius, 1977). In addition, the SW review made much of the view expressed by Olds and others that effectiveness is more likely when proximal (e.g. home support) and distal (e.g. health and community services) influences work synergistically, a view closely related to an ecological model of health promotion.
- Evidence in relation to Triple P (not provided in the initial SW document) supports the developers' contention that it 'has the strongest empirical support of any intervention with children' (Sanders, 1999, p.72). However, its evaluation has almost entirely been based on RCTs involving children and families with identified problems, such as maternal depression or conduct problems in the child, either receiving or in need of treatment. The extent to which the effects are generalisable to all children, and identifiable at the community level, is much less certain.
- While there is some evidence that effects diminish with time (Gutelius, 1977; Kendrick et al., 2000), other studies demonstrate mixed results, and some point to effects that persist or emerge later. In the Elmira trial, although no significant differences in HOME scores were found between intervention and control groups at 34 and 46 months (Olds et al., (1986; 1994), mothers in the intervention groups had fewer subsequent pregnancies and at age 15 their children had lower rates of substance use, arrests and convictions (Olds et al.,

1997). These findings raise the possibility of 'sleeper' effects (only emerging over time) though none of the RCTs, nor the systematic reviews, mention this.

- Almost all studies were and remain US-based, raising the question of transferability to other national/cultural contexts. This issue was mentioned in the SW review, which noted that an exact replication in the UK was unlikely because existing health visiting services routinely do some of the work of (intervention) nurses. The implication of this (which was not commented on) is that any intervention effect of enhanced health visiting is likely to be less in the UK than in the US.
- While several of the intervention studies included minority ethnic groups, none specifically addressed the question of the extent to which programmes were either applicable or effective in those groups. In the SW review, which gave prominence to the needs of Asian families, no studies of this ethnic minority were referred to.

In summary, the evidence on the effectiveness of home intervention programmes which was available to the SW team at the start of the project was less conclusive than often supposed. Many of the studies had methodological flaws, effect sizes were generally small and often confined to the most vulnerable of sub-groups. The best evidence for an intervention effect came from the most intensive, therapeutically oriented programmes of long duration.

Since then, further systematic reviews (Kendrick et al., 2000; Elkan et al., 2000) have highlighted methodological problems and urged caution in extrapolating beyond the US. Nevertheless, the results have been moderately encouraging. For example, in Kendrick et al's (2000) meta-analysis of 12/17 trials, a significant effect of home visiting on the quality of the home environment (as measured by the HOME score) was found, and some significant effects on other dimensions of parenting were also demonstrated in 20/27 studies. The authors noted, however, that their conclusion rested on an assessment of statistical significance rather than (as preferred) effect size. They also noted that since only 4 of 27 studies were from the UK 'caution must be exercised in extrapolating the results to current UK health visiting practice' (p. 443). Very recently, the Health Development Agency (HDA) (2004), have conducted a review of reviews of home-visiting programmes (including Kendrick et al). Their conclusion is equally cautious, the evidence being inconclusive in relation to several child outcomes (e.g. child abuse), though more positive for others (e.g. childhood injury). There is also some good evidence of positive effects on various dimensions of parenting or mother-child interaction.

2.1.3 Differences in design

There are a number of ways in which SW differed from interventions of proven effectiveness: these differences might be expected to affect its capacity to demonstrate an effect in Scotland. The most important of these are:

- **The setting** – Recognised throughout, the most obvious difference resides in the different health care systems in the US and UK, the latter having a system of statutory health visiting which routinely provides some of the features of intensive home visiting. Intervention effects in Glasgow are, therefore, likely to be smaller than those observed in the US trials. There are in addition cultural differences relating to the acceptability of such interventions which might impact on uptake, motivation and compliance, and hence effect.
- **The target population** – The SW target population differs from the trials. Rather than randomising families to intervention and control arms, it targets whole communities, so that intervention effects must be detected through comparison with a control community.

This reflects the underlying philosophy of the project, particularly the adoption of an ecological model, but there is no direct evidence of such a community-wide effect. Given that several (US) studies have demonstrated effects only on particularly vulnerable groups, it could be that the comparison of communities will obscure important effects in particular sub-groups. In principle this issue can be addressed by sub-group analysis, but only if the analyses are specified in advance and, where necessary, built into the design.

- **Ethnic minorities** – One of the SW areas (Glasgow South) includes a relatively high proportion of families of South-Asian origin, reflecting the prominence accorded to ethnic minorities in the proposal. While this may be desirable for a variety of reasons, there is no evidence supporting the effectiveness of home-visiting in this group, nor is it known how appropriate Triple P is in this context. Given the cultural differences in family structure and parenting, it is also unclear how applicable the key HOME measure is for this group. It is possible, therefore, that ethnic minority families will be a separate case, complicating between community comparisons.
- **The target child** – SW targets all children born to mothers in the two intervention areas while almost all the trials focus on first born children. This difference is not commented on in any document and it is neither clear whether this is likely to have any effect, nor what direction this is likely to be in. It is, however, possible that an effect would be more pronounced in first-born than subsequent children.
- **The intervention** – SW involves a dual intervention strategy combining home visiting with strengthened community support. In respect of both, the SW design is a departure from that of the interventions providing the evidence base.
  (a) Home Visiting – while SW contains many features of the effective interventions (frequent visiting with a specified educational and therapeutic content) it is clear that only a small proportion (19%) of families were visited in the ante-natal period. The best available evidence suggests that contact before the baby's birth is important. Reduced contact in this period might weaken an intervention effect.
  (b) Triple P - although regarded as the most evidence-based of all home intervention programmes, most of the evidence about effectiveness refers to specific sub-groups, notably children with conduct problems. As widely acknowledged, it is also not validated in a Scottish population, nor with respect to ethnic minorities such as those of South Asian origins.
  (c) Community support – although recognised as likely to enhance the effectiveness of home visiting, there is no direct evidence from RCTs on this issue.

The SW demonstration programme, therefore, had several characteristics which distinguished it from previous effective interventions which meant that there were departures from the evidence-base template. Although there were good reasons for this, it highlights the innovative nature of SW and further underscores the importance of its evaluation. As outlined in the introduction, the DPs were required to both utilise the evidence-base and to be creative. The key therefore is explicit clarity over what is evidence-based, what is new and what is an application to a new context.

2.2 Healthy Respect (HR)

2.2.1 The demonstration project

HR is a multi-strand programme (made up of 12 projects) to promote positive sexual health and relationships among young people. It has two overarching aims (1) to reduce teenage pregnancies and prevent the spread of Sexually Transmitted Infections (STIs) (2) to communicate learning about 'how and why outcomes emerge' (original bid, p.2) to promote

transferability. As originally specified, within these 2 aims, a number of specific objectives were outlined, including (Aim 1) promoting openness in discussion about sexual health, nurturance of self-esteem and responsibility and discouragement of coercive sex, and (Aim 2) the development of an HR website. HR, though predominantly incorporating existing organisations in Lothian, also developed innovative projects. Project B, Looked After and Accommodated Young People (LAAYP) was the only project that existed before the establishment of HR. Following the original bid (November 2000), which did not identify specific projects, a revised submission (project plan, July 2001) identified 12 projects (Projects A-L) and 19 reporting components. These projects referred to a range of existing and projected initiatives located in diverse settings, including improvements in the uptake of contraception post termination (A) and Chlamydia screening among women presenting for emergency contraception (D) in family planning clinics, provision of postal testing kits for Chlamydia (C) via several outlets, sexual health promotion in Further Education colleges (E) and schools (G), contact and support for specific groups such as LAAYP, (B) and Lesbian, Gay, Bisexual and Transgender Youth (F) and parents (I) and a project focussing on media campaigns (L). In combination, these projects comprise HR.

2.2.2 The evidence base

In the HR proposal, there was no separate review of the evidence on interventions to improve sexual health amongst young people. The only forms of 'evidence' presented were descriptive, with references to official reports highlighting the importance of sexual health education, self-esteem, multi-agency working and the role of poverty and a series of tables/figures detailing the sexual health profile of young people in Scotland, and where possible Lothian. For example, data from the Health Behaviours of Scottish School Children (HBSC) Scotland survey (Todd et al., 1999) was used to demonstrate the increasing proportion of 15 year-olds with sexual experience from 1990 to 1998, and the proportion in 1998 not using contraception at last intercourse. The implicit assumption throughout the HR proposal was that these Lothian data 'speak for themselves' and were simply further testimony to the accepted wisdom that the sexual health of young people needs to be improved. The following points are noteworthy:

- Except in the most general terms, the HR bid was characterised by the absence of an explicit evidence-base. The evidence adduced in support of HR was general, patchy and possibly selective. Much is made of the quality of relationships, and particularly the importance of self-esteem, but no supporting references were provided. While criticism of the self-esteem thesis in relation to risk behaviours is recent, some evidence on this issue was available at the time of the proposal (West and Sweeting, 1997).
- Most importantly, despite the fact that HR is an intervention (albeit multi-faceted), no reference was made to any intervention study in sexual health which might provide a rationale for the effectiveness either of the whole programme or any of its components. For example, there was no mention of the extensive literature on sex education in schools, (e.g. Wellings et al., 1995), nor was there any reference to a number of systematic reviews on (mainly US) sexual health education programmes. Although these systematic reviews identify several studies showing no intervention effects, on balance they provide evidence for the effectiveness of certain programmes in reducing certain outcomes such as underage sex and/or increased condom use (Kirby et al., 1004; Kirby, 1999; NHS Centre for Reviews, 1997). It can only be assumed that this literature was either known to HR but not made explicit, or that it was entirely taken for granted that sexual health education was a good thing regardless of its effectiveness.

- HR has as one of its goals the promotion of openness in sexual health matters. In support of this, reference was made to the low teenage pregnancy rate in the Netherlands, together with the view that this results from a positive and open sexual health culture. Although this is an admirable goal, there was no explicit acknowledgement of any problem of transferability to the Scottish context.

For all the DPs, the evidence base is constantly developing, and needs to be updated as new studies and reviews are published. In the case of HR, the evidence base was unacknowledged and possibly unknown to the bidders, although at a later stage in the bidding process (some time between January and April 2001), at the request of the Scottish Executive, an outline of the evidence base was prepared which outlined the Healthy Respect component project, its rationale and the evidence base for the intervention. Since its inception important new evidence has emerged which challenges parts of its underlying rationale. The most important new study is the MRC SHARE trial which provided the model for the sexual health intervention in the HR schools project (G) and the model for its evaluation. Disappointingly, the results of the trial (published in 2002, but made available to HR in July 2000), did not show an effect of the intervention on reported age of first intercourse or condom use, two of HR's key targets (Wight et al., 2002). However, consistent with other trials (Mellianby et al., 1995), improvements in knowledge were demonstrated, and it remains possible that differences in behavioural outcomes will be found in later follow-ups and by reference to termination rates. While differences between the HR project and the MRC intervention (HR has established links with health services and developed multi-disciplinary working) might be expected to increase an intervention effect, the evidence from the MRC study clearly runs counter to the expectations of HR and challenges the rationale. This raises the question of how a DP, or its sponsor, should react when contrary evidence emerges after the project is underway. In this case, the new evidence appears to have made no difference since the results of the SHARE trial were known to HR when the project bids were being developed.

2.2.3 Differences in design

Unlike SW, without an explicit evidence base for HR, there is no effective intervention template against which to assess the design of HR. HR proceeded as if it were an entirely new project, again highlighting the need for evaluation.

2.3 Have a Heart Paisley (HAHP)

2.3.1 The demonstration project

HaHP is an area-based, multi-component, multi-agency project with a strong community focus that aims to prevent coronary heart disease (CHD), to promote good health and to reduce health inequalities in Paisley, a town with high levels of CHD and of deprivation.

The original proposal set out a framework for a wide range of health promoting activities targeting the main risk factors for CHD and encompassing both primary and secondary prevention. The NHS, local authorities, voluntary groups and community groups are all partners in the development and implementation of the project. The HAHP framework encompassed both primary and secondary prevention of CHD. Four geographical locality networks provided the framework within which risk factor modifying activities are supported, co-ordinated and targeted. A fifth network co-ordinated the secondary prevention work aimed at people in high risk groups such as those with established CHD. Marketing activities aimed to raise awareness of the HaHP activities and brand and to endorse positive lifestyle messages.

A few defined initiatives were included in the initial proposal, such as the development of CHD registers as a basis for improved secondary prevention. However most of the component parts of the project were not defined in detail. The partner agencies were invited to submit specific sub-projects for funding from HaHP – focussing on the three main risk factors: smoking, diet and physical activity. Training needs were also seen as a priority issue.

2.3.2 The evidence base

Since the 1970s, many community based or population based intervention studies have been carried out for the prevention of CHD (Puska et al., 1995; Ebrahim and Davey Smith, 2001; Farquhar et al., 1990; Luepker et al., 1994). Although the original proposal for HaHP included an appendix on the evidence base, this was mainly devoted to describing the North Karelia study, with some other examples of prevention initiatives. The proposal did not include a comprehensive and critical appraisal of the admittedly large and complex literature on community based CHD prevention studies or on risk factor based interventions. This would have been impossible within the time available to the bidders. However a good summary was included as part of the successful bid to evaluate HaHP (Hanlon et al., 2000) and later on a detailed review was conducted as part of the evaluation (Blamey, A., 2002).

- **North Karelia Project**

This was set up in Finland in the 1970s (Puska et al., 1985; Vartiainen et al., 1994) and is the best known of the many community based CHD prevention initiatives. It has been extensively evaluated, through a series of population surveys and other methods (Puska, 1995). The project built on the support of local community leaders and the general public and was intended to provide a unified and comprehensive approach to CHD prevention. It involved the media, workplaces, primary care, hospitals, schools and local communities. A very wide range of initiatives were implemented, including training programmes, mobilisation of public support through local leaders, formation of housewives groups and targeting of grocery shops and the food industry.

Over the duration of the study, there has been a substantial fall in CHD mortality rates for CHD in North Karelia and in levels of risk factors. However, a similar downward trend is seen throughout all parts of Finland, raising the hotly debated question of how far the mortality decline is attributable to the specific interventions in the programme (Ebrahim and Davey Smith, 2001).

- **Other community based interventions**

In the 1980s several other influential community based studies were conducted in the USA, such as the Stanford Heart Disease Prevention Programme, Stanford Five Cities Report, Minnesota Heart Health Programme and Pawtucket Heart Health Programme (Farquhar et al., 1990; Luepker et al., 1994; Carleton et al., 1995). In each case, the differences between the intervention and comparison areas in risk factor modification were small or non-significant and the attributable effect on mortality was also disappointing. For example, a report from the Stanford Five Cities project found a similar decline in disease rates in intervention and control cities, suggesting that influences outwith the programme accounted for the observed changes.

Within the UK, a community based demonstration programme in Wales, aimed at reducing risk factors for CHD, was set up in the 1980s. Heartbeat Wales drew on the experiences of

the studies in Finland and the United States. While a reduction in reported smoking prevalence and improvements in dietary choice were observed, there was no net intervention effect for the programme areas over and above the observed change in the control area of NE England (Tudor-Smith et al., 1998). The investigators concluded that 'with hindsight, the difficulties of evaluating such a complex multifaceted intervention were underestimated. Further debate on the most appropriate methods for assessing the effectiveness of community based health promotion programmes is called for.'

- **Multiple risk factor intervention studies**

In parallel with these community based studies, there have been several large RCTs of multiple risk factor interventions in reducing CHD mortality. Ebrahim and Davey Smith reported a systematic review and meta-analysis of RCTs in workforces and in primary care in which individuals were randomly allocated to more than one of six interventions (stopping smoking, exercise, dietary advice, weight control, antihypertensive drugs and cholesterol lowering drugs). The changes in risk factors were modest and related to the amount of drug treatment used. Interventions using education, with or without drugs, were more effective in people with hypertension and in other high-risk groups. (Ebrahim S., Davey Smith G., 1997).

- **Secondary prevention**

In contrast to the equivocal evidence on community based programmes for primary intervention, there is a wealth of good evidence from RCTs and observational studies about the effectiveness of secondary prevention of CHD - in people who already have signs or symptoms of the disease. In addition to demonstrated evidence about the effectiveness of individual interventions there is also evidence from a systematic review that multidisciplinary disease management programmes improve the processes of care, reduce admissions and enhance quality of life in patients with CHD, although the impact on survival and recurrent heart attacks remains uncertain (McAlister et al., 2001). Much of this evidence was available to the HaHP team at the start of the programme and developing a systematic approach to implementing evidence-based guidelines was a major component of the secondary prevention project.

In summary, although some components of HaHP, notably the secondary prevention initiatives have an evidence base, evidence for the effectiveness of community based prevention programmes is, at best, equivocal. It certainly does not consistently demonstrate additional benefit in accelerating the background secular trends in health behaviour and in CHD mortality reduction which are apparent in Western countries.

2.3.3 Differences in design

Given the lack of a good evidence base for HaHP as a whole, the design of the demonstration project only compounds the difficulties. The problems in designing and implementing an effective evaluation of population based projects like this have previously been highlighted (Tudor-Smith et al., 1998; Puska, 2000).

The HaHP project has also taken place against a background of falling CHD mortality rates in Scotland and there is little realistic chance of demonstrating additional benefits in accelerating these trends. The secondary prevention elements had more potential to demonstrate effectiveness and to highlight useful lessons for other parts of Scotland. In addition to this lack of a strong evidence base for community-wide CHD interventions as a whole, there are a number of other ways in which the design of an effective evaluation of the HaHP project was problematic. The most important of these are:

- **Interventions**

Both the intensity and the quality of the interventions used in most community based intervention studies are difficult to measure and this was particularly true of HaHP.

Many small projects were funded with variable levels of activity, both over time and throughout the geographical area. This made it difficult to identify which parts of the project might have an effect. Many of these specific interventions were not evidence based.

There was anecdotal evidence that, prior to HaHP, levels of some health promoting interventions such as smoking cessation support, were lower in Paisley than elsewhere in Scotland. This could tend to exaggerate any benefit for HaHP.

In addition, HaHP took place against a background of many national initiatives on improving health in Scotland. These would have an impact on the population of Paisley and would make it very difficult to isolate any specific benefits from HaHP.

- **Setting**

Although mortality from CHD in Scotland remains amongst the highest in developed countries, death rates have fallen substantially over the past 25 years and this trend is continuing throughout the whole country. The population of Paisley has high levels of deprivation and this is associated with high levels of CHD. There are significant cultural differences between the intervention area and North Karelia which are likely to impact on uptake, motivation compliance and hence effectiveness.

2.3.4 Summary

A major problem for HaHP is that community based CHD prevention studies have failed to show a clearly attributable effect, possibly even in the case of North Karelia. HaHP took place against a background of falling CHD mortality rates in Scotland. It had little realistic chance of demonstrating additional benefits in accelerating these trends.

The secondary prevention elements had more potential to demonstrate effectiveness and to highlight useful lessons for other parts of Scotland.

Section 3: Evaluation

In this section we first consider some general issues regarding the relationship between policy making and evaluation (3.1) then go on to describe the approaches adopted by the external evaluations of the demonstration projects (3.2), the internal monitoring and evaluation work (3.3) and finally consider the relationship between the two (3.4).

3.1 Approaches to evaluation

Evaluation is a form of applied research concerned with assessing the results, impacts and/or outcomes achieved by some form of intervention (whether this be a project, a programme, an institution or a policy) in order to inform judgements about that intervention. While there are many different types of evaluation and methodologies, essentially there are two broad approaches:

1. Evaluations that are concerned with proving effectiveness; they are concerned with the achievement of aims/objectives and impact/outcomes and with explaining success/failure; these are analysis-oriented and framed within agendas concerned with accountability and knowledge-building
2. Evaluations that are concerned with improving the implementation of a programme or policy, or strengthening institutions, communities or networks; these are action-oriented and often framed within agendas concerned with development and empowerment.
   (Chelimsky, 2001; Stern, 2004)

Combining these two approaches within a single evaluation can lead to tensions since not only will the basic purpose of the evaluation differ, but there are also differences in the epistemological and methodological approaches adopted, and the relationships between the evaluators and those involved in programme implementation (independent observation and analysis versus active engagement).

One major consideration in the selection of appropriate evaluation approaches and designs is the stage of the programme or policy cycle – planning, development, implementation. The alignment of the evaluation focus with the stage in the policy/programme cycle is formalised within many evaluation frameworks. For example, the European Commission differentiates between ex ante evaluation (sometimes called 'appraisal') and impact assessment, which are undertaken at the development stage, and interim and ex post evaluation, which are undertaken once the intervention has been implemented (European Commission, 2003). Within the current UK (evidence-based) policy context, and certainly within public health/health improvement, there is also an emphasis on the use of systematic reviews of results from evaluation studies to inform the planning stage in the programme/policy cycle and to ensure that proposals are evidence-based. This is formalised within the HEBS (now NHS Health Scotland) evaluation framework for health promotion (HEBS, 1999; Wimbush & Watson, 2000).

Implementing this kind of systematic, staged approach to evaluation can be difficult. Many government programmes are funded with the expectation of immediate implementation and delivery within a fixed, e.g. three year period, with little allowance for an initial developmental/design phase. One implication is that evaluations of programme effectiveness are commissioned at too early a stage. There is seldom consideration of the need for an initial phase of formative evaluation while the programme is being developed and set up so that

interventions can be piloted for their feasibility, acceptability and likely effectiveness within a particular local context, and thereafter adjusted to improve their chances of effectiveness. This exactly describes the situation in which the DPs, and their evaluations, developed.

3.2 External evaluation

Among the bids for external evaluation, two contrasting models were proposed. One of these, stemming from the clinical trial and health services research tradition, employed an objective, 'hands-off' approach to evaluate the outcomes from HR, while the other employed a much more collaborative developmental and 'hands-on' approach, using Theories of Change (ToC), to first clarify the objectives of the SW and HaHP projects and then monitor progress towards these objectives. All evaluations used a quasi-experimental research design as one component.

The adoption of contrasting approaches to evaluation permits an assessment of the strengths and weaknesses of each approach and some of the problems from the viewpoints of both the DPs and the evaluation teams in adopting one model rather than another.

3.2.1 Healthy Respect: evaluation design

In preparing their evaluation of HR, the external team (HR/E) included the top line aims and objectives of HR to improve sexual health for young people in Lothian, each generating a number of 'pre-specified' hypotheses. The evaluation focussed on (a) sexual health outcomes of young people in Lothian by reference both to routine sexual health data (e.g. teenage conception and termination rates) and survey data on secondary school pupils' sexual health knowledge, behaviour and uptake of services (b) the organisation and performance of interagency partners in the provision of sexual health services and (c) the implementation and process of each HR component project. The evaluation would test the following hypotheses relating to the first two objectives (a) HR would impact on attitudes and behavioural change (better communication with parents/teachers on sexual health issues, reduction in proportion having underage sex, and increased knowledge and (reported) use of condoms), service access, acceptability and uptake; reduce conception/abortion rates and increase rates of Chlamydia testing (b) HR would increase interaction and networking between service providers to the perceived benefit of clients. The third objective (c) focussed on the processes by which these changes were hypothesised to occur.

To address these objectives/hypotheses, HR/E proposed the following evaluation to be conducted over 4 years from November 2000 to November 2004: (a) a quasi-experimental 'before-after design' comparing young people in the intervention area (Lothian) with a comparison area (Grampian), selected both for practical reasons and to represent another East coast region (thus avoiding west of Scotland cultural and religious complications), to be supplemented by qualitative data derived from focus groups. In the case of school surveys, representative samples of S3/S4 pupils in each region would be selected and power calculations (based on responses to the Scottish WHO HBSC [Todd et al., 1999]) indicated that around 2000 boys and 2000 girls were required in both intervention and comparison areas to demonstrate an effect (e.g. on underage sex); (b) a mapping exercise of sexual health and related services combined with interviews with agency personnel. To address objective (c), HR/E proposed qualitative (e.g. interviews and focus groups with clients and service providers) methods, each designed to identify process measures and implementation of the individual projects. HR/E were clear from the outset that with multiple concurrent initiatives, it would not be possible to attribute an intervention effect on population-level outcomes to individual components of HR.

3.2.2 Healthy Respect: a 'hands-off' approach to evaluation

The underlying model of the HR evaluation was therefore based on the assumption of an intervention with fixed overall HR aims and objectives, from which specific hypotheses could be formulated and which were testable by reference to a design combining a quasi-experimental method to identify an intervention effect on specified outcomes, with various qualitative methods used to illuminate processes which might bring this about. In this model, the independence of the evaluators is imperative with contact between the evaluators and implementers kept to a minimum for data collection purposes. The evaluation team do not seek to influence the direction or development of the intervention, nor is the evaluation open to influence by the project team. While consensual and collaborative working with the HR team is necessary to obtain data, the relationship is otherwise non-interactive. In the revised proposal (August 2000), HR/E specifically identified that one of their goals was to 'avoid contamination of HR' (p.6). Thus, HR/E have not fed back interim results to HR on the grounds that this might influence the direction of the project. An alternative approach such as theory of change was described as being outwith the approach to HR evaluation.

The adoption of the 'hands-off' approach to evaluation used by HR/E seeks to provide an unbiased test of the initial hypotheses. However, there are potential disadvantages:

- It assumes that the demonstration and component projects' objectives are stable. Any change in objectives, or failure to implement them, makes the project less evaluable. Although identification of such changes is one of the reasons for the qualitative process evaluation and provides evidence about the extent to which the intervention was implemented according to plan, the shifting emphasis of HR towards process rather than outcome measures was identified as a major problem for HR/E. It was perceived as a departure from their original understanding of HR's preparedness to be tested for impact and effectiveness. (4.11.03).
- It assumes objectives are clearly articulated and communicated. This was not the case in the early stages of HR, both the project objectives and management continuity being unstable for at least 8 months from August 2000. Against this background, HR/E drew up a Memorandum of Agreement to agree objectives and mechanisms for process evaluation, roles and responsibilities and respective intellectual property right areas. The Memorandum was consistent with the commissioned evaluation (August 2000) in not offering interim feedback
- The lack of feedback from the evaluation to HR was experienced as frustrating by project staff, leaving them feeling they lacked guidance and any sense of whether or not they were achieving their objectives.
- The 'hands-off' model of evaluation also runs the risk of generating a gap in expectations between internal and external evaluation teams. In the early stages of HR, the perception that HR/E had not given sufficient attention to self-esteem led the project manager to develop a separate research proposal to address the issue, which was not subsequently funded. There were many other examples of HR component projects undertaking research (often of dubious quality) to fit in with a culture of evaluation internal to HR. One consequence of this is that internal and external evaluations may produce disparate process findings.

3.2.3 Healthy Respect: assessment

The external evaluation involves several components including health outcomes of teenage pregnancy, compliance with national recommendations for the detection and management of Chlamydia in the context of the National Chlamydia SIGN Guideline Audit and service provision for STIs. Central to it is a repeat cross-sectional survey of sexual health knowledge, attitudes and (reported) behaviours among S3/4 secondary pupils in intervention and control schools (Lothian and Grampian respectively). The first survey, involving 10 Lothian (2760 pupils – 80%) and 5 Grampian (1501 pupils – 83%) schools, was completed between September and December 2001, the second in 9/10 Lothian and all 5 Grampian schools completed during the same period in 2003.

In principle, this design should be able to address the key hypotheses of the evaluation even allowing for the risk of contamination between areas, and the difficulty of causal attribution of effects to the intervention given the multi-faceted nature of HR. The surveys appear to have been conducted efficiently with good response rates within schools (although absentees were not followed up). However, in several respects the design was less than optimal:

- The original HR/E objective (August 2000) to compare representative samples of schools (and pupils) was not achieved, and in the case of Lothian was not achievable because HR was only operating in selected Lothian schools. The ten intervention schools that 'signed up' to HR were self-selected, i.e. were 'volunteers'. They may represent those most committed to sexual health education, potentially biasing estimates of the intervention effect.
- The original aim for the control sample was to identify ten schools in Grampian matched by size, rurality and level of deprivation. In the event, only five of the 17 selected and invited secondary schools in Grampian, agreed to take part. These schools may not be representative of Grampian schools, nor were they well-matched controls for Lothian schools. This may compensate for the bias in the Lothian sample, but the degree of under or over compensation is impossible to estimate, making interpretation of results difficult and generalisation risky. The potential selection bias and hence representativeness of the schools is an issue that will be addressed explicitly in the final report.
- The reduction in sample size caused by school recruitment problems necessitated a re-assessment of the power of the sample to detect specified effect sizes. The original estimates (above) were 2000 pupils of each sex in both intervention and control areas, the response to the first survey indicating this was not achieved, especially in Grampian. The Progress Report (October 2002) notes, however, that the achieved sample sizes of the first survey based on the revised sample estimate (using a new 2:1 Lothian/Grampian ratio), conducted before the first survey, 'continued to allow detection of an effect size of 4% to 5%' (p.3). This was based on the revised power calculation agreed with the CSO, assuming a 4-5% difference in prevalence of an outcome such as reported experience of sexual intercourse given a baseline prevalence of approximately 20%.

We do not yet have any results from HR/E on the school-based intervention. It seems likely, however, that the problems of school recruitment may make these results difficult to interpret. If there is no difference between Lothian and Grampian pupils (controlling for confounders), it could be attributable to a number of factors. If there is, it may not be generalisable to a wider population because of potential biases introduced by the schools selected.

3.2.4 Healthy Respect: process evaluation

Process evaluation was a limited part of the original research proposal from HR/E, but was extended in the revised version at the request of the Scottish Executive. It includes:

- Assessing the effectiveness of interagency working using descriptive before-and-after inventories and mapping of service provision, partnerships and networks and through observation of professionals' contacts and activities through diary keeping.
- Describing the implementation process of HR's 12 component projects by identifying key process indicators of implementation for each of the projects. The views of projects' clients and providers are sought to identify best practice, perceived impact and acceptability.

The process evaluation was intended to be used to identify, understand and interpret any observed changes in outcome measures, making associations between process and outcome where strict attribution is not possible. While the process evaluation and context mapping helped keep the evaluation team up to date with the evolution of the demonstration project, the team sought to maintain their independent position by avoiding feedback from the process evaluation.

So far, at the time of writing, there have been no reports from the process evaluation of HR. The contribution of this element of the Healthy Respect evaluation to understanding causal attribution is likely to be weak without an overall programme theory to make the links between goals, individual project activities and outcomes. The process evaluation will only aid in assessing the effectiveness of each component project in terms of their own objectives, against a set of 7 criteria derived from literature on good practice on health promotion.

3.2.5 Starting Well: evaluation design

3.2.5.1 Theory of change – a 'hands-on' approach to evaluation

In the (revised) bid to evaluate SW, the external team (SW/E) made a distinction between the criteria and methodology used to evaluate an intervention trial and those appropriate for a 'demonstration project', the rationale for the latter being as much about 'improving the intervention as proving that it works' (p.7). The idea that the evaluation should shape the direction of the intervention contrasts with the approach described above. The rationale for this more interactive approach is a recognition that 'real-life' interventions rarely stand still and often depart from their initial objectives, as a consequence of external events, such as policy changes or service reorganisation, or internal changes in direction initiated by those implementing the intervention. Indeed, initial objectives themselves may not be clear. From this perspective, involvement of the evaluator in the development and course of the DP is desirable since it provides a means of identifying what is being evaluated. The method proposed to address these issues in the SW (and HAHP – see below) evaluations was the 'theory of change' (ToC) (Fullbright-Anderson, 1998; Judge & Bauld, 2001).

ToC seeks a better understanding of the processes in an intervention that might produce predicted change. The first step is to identify the connections made by key stakeholders between DP inputs and desired outcomes to make an assessment of the likelihood that the goals can be achieved. Thus, in this first stage, through interaction between SW and SW/E (interviews/focus groups with steering group members), the aim was to clarify objectives. After the initial stage, the focus of ToC switches to the documentation of processes designed to assess whether intended actions take place and whether predicted changes are observed.

3.2.5.2 Process Evaluation

In the case of SW/E, the process evaluation identified three key issues that were an integral part of SW's TOC: the extent to which intensive home visiting led to the development of therapeutic alliances between families and their home visitors; the implementation issues involved in developing a skill mix approach to home visiting; and the degree to which intensive home visiting at an individual family level led to improved community and strategic responses to child and family health problems. This took the form of detailed case studies with 59 SW families and associated Health Visitors (HVs) and lay support workers in order to evaluate the extent to which specified components of the home intervention (e.g. the Family Health Plan) were being achieved. It also involved detailed documentation of what actually happened during the implementation of SW in respect of all its components. In a later document (Shute & Judge, in press), these were reduced to 3 central components (a) case studies (as before) (b) formation and development of staff team of professionals and paraprofessionals (c) influence of identified health needs on local and higher-level planning. While these methods and measures would normally be part of the process evaluation in any intervention, the 'theory of change' involves the systematic documentation of change in intervention components as they relate to intended outcomes. It is claimed that this facilitates a more sensitive analysis of 'causal' attribution than is often the case, enabling better identification of both the reasons for intervention success and failure.

3.2.5.3 Quasi-experimental study

The use of ToC, however, was intended to complement rather than replace a traditional quasi-experimental approach. SW/E proposed a comparison (not unlike that of HR/E) of two cohorts, one in the SW (intervention) area, the other in a comparison area with broadly comparable demographic and socio-economic characteristics. The SW/E cohort is a time-specified subset of all SW participants, defined in this (and the comparison cohort) as all babies born between June 2001 and June 2002 . It was initially proposed to recruit families via HVs at the point at which they first made contact (assumed to be in the antenatal period) and follow them up for a period of 30 months (contacts at birth, 6 weeks, 6 months, 18 months and 30 months). However the latter follow up was abandoned. Thus, in addition to implementing the SW programme, both recruitment to and administration of SW/E instruments would be conducted by HVs, supplemented by SW/E interviewers in the comparison area. In general, this was what happened, recruitment and some instruments (e.g. postal questionnaires) being administered by HVs, trained research nurses additionally conducting home interviews at 6 and 18 months, including the administration of the HOME measure.

A range of outcome measures was proposed for each point of contact including maternal and family characteristics (antenatal), birthweight/gestation (birth), postnatal depression (6 weeks), maternal diet, breastfeeding, immunisations, HOME score (6 and 18 months) etc. Reflecting the emphasis on parenting as a key outcome, the HOME score (with 6 sub-scores) was seen as the core outcome measure. The projected number of families in the SW area was 600-700 per year (1500-2000 over 3 years). Two options were given for the comparison area involving (a) a one year or (b) three year cohort sample respectively. Power calculations were given in relation to predicted effect sizes on HOME and child accidents, the latter being used to demonstrate that in option (a) the sample size would not be adequate.

3.2.6 Starting Well: assessment

3.2.6.1 Theory of Change

ToC is an interesting development in evaluation methodology and seems particularly appropriate when the initial objectives of a project are unclear. However, the following points are worth noting:

- ToC has as one of its overall aims the improvement of interventions. It does not start with objectives as defined by a project (as with HR/E) but seeks via interaction with DPs to clarify objectives and possibly redefine them. This is a departure from the usual scientific principle of independence and might in various ways compromise the implementation of the DP. For example, DPs might become over-dependent on evaluators for direction and guidance. This is unlikely to be a problem in the formative stages of a project when objectives are being formulated, but if objectives (and related processes) continue to change in a 'mature' project, it becomes more difficult to know what is being evaluated. ToC would, however, facilitate an understanding of how and why such change occurred (e.g. the DP became committed to a new approach), but what then is the 'project' that is being evaluated overall?
- Inasmuch as ToC leads to a change in objectives, there is a risk that the DP shifts away from the evidence base it rested on in the first place, thereby inadvertently reducing the likelihood of obtaining an intervention effect
- Any change in project objectives resulting from ToC that occurred after an evaluation (e.g. before/after survey) had been commenced would potentially render that evaluation flawed. This highlights the fact that ToC is of particular value in the formative stages of a project and should not intervene after a project is up and running. Beyond the initial stages, ToC functions much like any other process measure (a point made by HR/E).
- ToC may place demands on DPs to become involved in procedures they may not have expected and might not want. The view from HaHP would have been that this was another set of controls while SW did not regard the ToC feedback as particularly useful.

Since both SW and HaHP process evaluations used ToC methods, a combined discussion of how well ToC worked in practice is included in section 3.2.8 below.

3.2.6.2 Quasi-experimental study

As in the case of HR/E, the choice of a quasi-experimental design was an appropriate methodology to test for differences in outcomes between intervention and comparison areas. There were some modifications to the design resulting both from delays in implementing SW and from the practicalities of conducting all the proposed follow-ups within the designated timeframe. Thus, the number of SW/E contacts has been limited to three points (10-14 days after birth, 6 months and 18 months) and the availability of data has been limited by the way SW was itself implemented (most notably in the lack of contact with families in the ante-natal period). However, while these problems restrict the capacity of the quasi-experimental design to deliver they do not invalidate it. The extent to which it has delivered depends on the following considerations:

- Recruitment to the survey was more problematic than anticipated. Although 98% of all families participated in SW, HVs were less successful in recruiting them to SW/E. In the SW areas, of 604 births in the specified time period, 375 (62.5%) mothers agreed to take part; in the comparison area, of (an estimated) 600 births, 262 (43.7%) consented to participate (total n=627), possibly reflecting a lower level of commitment among non-SW HVs. Consequently, recruitment of controls was later extended to health visiting teams in

the West of Glasgow. It is not clear how this affected the representativeness of the control sample.

- Response rates to each of the SW/E contacts have been lower than expected. In the first (baseline) contact, involving a postal questionnaire, data were only available on 447/637 (70%) of cases (71% SW area, 69% comparison area); at six months (for interview) the comparable figure was better 493/637 (77%) of cases (80% SW, 73% comparison). However, an analysis of six month outcomes (Shute and Judge, in press) was based on all those with baseline and six month data, reducing the sample to 359/637 (57%) or 30% of the population of families in both areas. The number of cases available for analysis (SW 213, Comparison, 146), was below the earlier estimate of the numbers (220 families per area) needed to detect an intervention effect, raising the possibility that relevant effects would fail to be detected.
- These problems raise the question whether each sample is representative of its respective population, previously regarded as 'a prerequisite for this evaluation aim' (First Annual report, p.5). This issue was not addressed in the paper analysing six month outcomes (Shute et al., submitted) where both areas were reported to have similar proportions of lone mothers and non home-owners to the Glasgow population. This does suggest the possibility that the SW/E samples were biased towards less deprived families in both the SW and comparison areas. The issue was highlighted in the 18 month progress report (e.g. within area comparisons of opt-ins and opt-outs) and should also be addressed in the final report.
- It was also evident from the six month data that outcomes for ethnic minorities differ (higher maternal depression, poorer HOME scores), which raises the question of the applicability of the SW intervention to this particular sub-group or the validity of these instruments in a transcultural context. Although this can be controlled for in multivariate analysis, a sensitivity analysis to determine how it impacts on the overall intervention effect would be useful.
- Throughout its development, SW has evolved, to the extent that it is no longer regarded as a project so much as an approach (Ross & de Caestecker, submitted). This has, as acknowledged, made it extremely difficult to evaluate since families have not been exposed to a constant intervention but rather to different types and levels of intervention. Thus, in the preliminary phase, the early families may not have received the full intervention, in later phases the intervention may have become diluted, the result of which is that without taking this into account it may not be possible to detect an effect of SW at its most optimal. This issue was alluded to by Shute & Judge who used the concept of the 'mature' DP to indicate the point at which a project is fully up and running and to some extent constant.

In summary, although SW/E have conducted this part of the evaluation to a high standard, problems of recruitment have limited its capacity to deliver. It is not yet clear how representative the samples are of their respective populations, nor whether the numbers are adequate to demonstrate an intervention effect on parenting. In retrospect, it may have been better to seek consent from mothers in the ante-natal period (with additional data collection benefits) and to collect baseline data via interview rather than postal questionnaire. It is also possible that the full effect of SW has been obscured by changes in the DP over time which might be revealed via identification, and related analysis, of a 'mature' phase.

3.2.7 HaHP: evaluation design

The external evaluation of HaHP involves 4 separate but linked approaches that are intended to give a balanced perspective on the overall processes, impacts and outcomes of the demonstration project.

These are:

- Theory of Change
- A quasi-experimental survey
- Contextual analysis
- A range of interrelated studies of key settings and organisations (the community, primary care and the local authority)

For a variety of reasons, planned and unplanned, both the HaHP project itself and the detailed format of the evaluation have developed and altered over the period of the project. For example, the interrelated studies were introduced to strengthen the evaluation once it became clear that problems with the surveys would limit their usefulness. The integrated case studies focused on two settings (primary care and community) and one organisation (local authority), and looked at the extent of service development and the impact of HaHP on professionals and/or agenda change, at both strategic and operational levels.

Many of the general points made in relation to SW/E also apply to the evaluation of HaHP (HaHP/E). Without the final report we cannot judge the success of the overall approach. Here we consider issues related to the quasi-experimental surveys.

3.2.7.1 Quasi-experimental study

This component of the evaluation aimed to assess the impact of the overall intervention in Paisley. A comparison area, Inverclyde, was identified, with similar population characteristics and geographically adjacent to the study area.

Randomly selected adults aged 20-70, within age and sex quotas, were to be assessed for CHD risk factors and health related behaviours at the beginning and reassessed at the end of the intervention period in both Paisley and Inverclyde. The assessments consisted of a questionnaire and attendance at a nurse led clinical assessment.

However, despite major efforts by the evaluators, the response rate for the baseline survey was a disappointing 28% in Paisley and 27% in Inverclyde. Changes in data protection regulations and inaccuracies in the addresses on the community health index caused particular difficulties which could not necessarily have been anticipated. The low response reflects recent experience in heath and lifestyle surveys throughout Scotland where many areas have suffered falling levels of response in the past few years.

The poor response rate and an over-representation of older people and less deprived areas meant that the survey population was not representative of the Paisley and Inverclyde populations. Several options to tackle this problem were considered and the evaluation proposal was revised as a result. The revised approach aims to use secondary data, including monitoring information from within the project.

This is likely to cause different problems and it may prove difficult to get good quality comprehensive data on trends in risk factors and health related behaviour in the general population and in various sub groups such as young people. This has obvious implications for

HaHP as a national demonstration project for CHD prevention where any impact of the interventions on mortality would not be seen for some years and some effect on intermediate measures such as risk factors would be expected.

The design of this part of the evaluation was problematic from the start, with most other community based CHD prevention studies being unable to demonstrate an attributable reduction in risk factors in the study areas compared to control areas. In summary, the complex nature of HaHP, the unrealistic timescales for planning and several unforeseen constraints have led to real problems both for the implementation and the evaluation.

3.2.8 Theory of Change in practice

The process evaluation element of the external evaluations of Starting Well and Have a Heart Paisley consisted of two distinct strands:

- A Theory of Change strand which attempted to explore, surface and develop the 'programme logic' (i.e. the logical connections between the programme's aims, the programme's activities and the intended outcomes) through interviews with key strategic and operational staff and the analysis of key documents
- A formative evaluation element that examined and described the implementation process of key elements of the DPs using qualitative research methods

Both these elements were intended to provide feedback and learning for the DPs themselves and for the wider policy and practice communities. Interim reports were produced in 2003 based on the early findings from these process evaluations elements (Mackenzie, 2003; Blamey, 2001 and 2003). The contribution of the ToC approach is said to be threefold (Weiss, 1995):

- Sharpening planning and implementation
- Facilitating the development of an evaluation framework
- Reducing problems associated with causal attribution

The extent to which these objectives were realised within HaHP and Starting Well was explicitly addressed in a paper (Mackenzie, M. and Blamey, A , 2004).

a) *Sharpening planning and implementation*. From the perspective of the HaHP and SW demonstration projects, ToC was 'a dominant element' of the external evaluation but it had mixed results. It was seen as an interesting and helpful tool in terms of project development and providing formative feedback to project management, giving them a mandate to make proposals for future developments. On the other hand it was also seen as burdensome and the feedback was sometimes too late to inform decisions. An alternative view was expressed by HaHP/E indicating that the early work in HaHP (Blamey 2001) made clear comment on the lack of evidence-base practice in key areas of HaHP as well as the over ambitious nature of the plans, yet these issues were not fully addressed by the project.

The ToC approach proved particularly useful for HaHP because of the major difficulties the DP faced in its early stages. The initial plans for HaHP were overly ambitious, and, as a result, timescales for delivery were lengthened and expectations of outputs and outcomes were reduced. There was a lack of initial planning time for such a complex initiative as well as operational problems such as staff recruitment and retention. (Blamey 2003). As a result,

the theory based approach was very well received by the HaHP project team, who appreciated the support and direction it enabled the team to develop.

For SW, the process of reflection involved was seen as the most useful aspect of the ToC approach. For both SW and HaHP, the main shortcoming was that it was a tool more appropriate to use at the planning stage, before the project started, rather than while the project was in development and maturing. From the perspective of the external evaluation teams, ToC was a helpful tool for surfacing conflicts in project goals, priorities and approaches among key stakeholders, but offered nothing in resolving these.

b) *Facilitating the development of an evaluation framework*. The ToC was intended to provide a framework of expectations for the evaluation to test. The value of ToC before the start of the projects is to clarify project components, linking them to projected outcomes and thus clarifying the fit between overall project objectives and overall project 'package'. From the projects' perspective, ToC can and did play a role in project design/clarification of projects elements/package as a whole, and guiding of evaluation. However, the implied sequential process was not apparent; it remained stuck in the project design phase rather than informing the evaluation focus. The main barrier for the projects to making the process work optimally was timing, it being desirable to apply ToC at design stage rather than once a project is up and running. From the evaluators' perspective, on the other hand, it was seen as useful in identifying key questions for the internal evaluation and to help the external evaluation team to further specify evaluation questions and prioritise the focus of the evaluation (eg roles within the workforce).

c) *Attribution - understanding what brought about the observed effects*
In order to fully realise the potential of the ToC approach in helping to unravel problems of causal attribution, it has to be carried out in a very intensive way to create a well specified, detailed theory of change. This was not possible to achieve in the context of SW and HaHP which are both complex multi-stranded interventions, developed and implemented in a fast-moving climate where detailed, highly specified planning does not exist and large areas of the projects are not grounded within an evidence base. A further reservation was the linearity of causal effect implied by the ToC approach, whereas in complex systems the synergistic effect of interaction between the individual components needs to be allowed for - a project may be ineffective on its own but be effective in the wider context of the overall programme.

3.3 Internal Monitoring and Evaluation

It was widely acknowledged by the DP teams that the establishment of an effective internal monitoring programme was given insufficient priority within the early stages of developing and implementing the projects. This led to delays in establishing proper performance monitoring systems, compounded by difficulties in recruiting evaluation staff with the appropriate skills and experience.

3.3.1 Healthy Respect

After a difficult beginning with several project management changes, it was only once the current project manager was in post that an internal evaluation function was developed. Standard project management methods have been adopted with an observational, process-focused approach. This involves using pro-formas to collect quarterly 'audit' information from all the constituent projects and providing feedback reports to projects for discussion. The data are analysed by the NHS Board's Health Information Unit and have been used to inform decision-making about continued funding of the projects. The continual demand for

monitoring data was seen as imposing a heavy burden on the smaller sub-projects. Schools have been reluctant to provide information on SHARE training coverage due to lack of capacity. Some additional research (eg media evaluation) has been commissioned by the internal evaluation officer.

After initial difficulties, the relationships between Healthy Respect's internal team and HR/E were described as 'good' in terms of collaboration and information sharing. The Memorandum of Agreement between HR and HR/E was seen as very important in delineating the respective roles of internal and external evaluation teams and avoiding duplication. The external evaluation team drafted a pro forma for the individual projects to use for their quarterly 'audit' returns (in lieu of diaries), but the external team have been reluctant to get involved in the projects and interim analyses of their survey data were strictly proscribed.

3.3.2 Starting Well

At the outset, there was little appreciation of the need for and role of the internal evaluation function. It was not until March 2002 (2 years from the start) that Starting Well was able to recruit someone with the appropriate skills for the internal evaluation post. The internal evaluation has produced a monthly management report using Family Health Plan data that is captured and collated centrally on a database in each area. Feeding the information back to the teams has served as a focus for team discussions. There has been an increasing demand for this information and ad hoc analyses have been conducted. The external evaluation team uses this internal monitoring data for the families in their cohort. Issues relating to data protection had to be sorted before data sharing was possible. Other research has also been commissioned by the internal evaluation officer once she was in post, such as action research on nursing practice, the use of practice guidelines and further work on community development.

3.3.3 Have A Heart Paisley

Similarly, the internal evaluation was severely delayed with HaHP. The project had expected the external evaluation to fulfil all requirements. Thus, the internal evaluation post was created at a junior officer level. Problems mounted due to limited baseline data collected at the start of the project, a low response rate to the baseline survey, problems in recruiting and retaining the internal evaluation post holder and a realisation that the external evaluation role did not encompass internal monitoring and evaluation functions. As the project evolved the interface between internal and external evaluation became more blurred, and a more integrated approach developed. Relationships between the HaHP project and the external evaluation team were described as good in terms of useful feedback and a constant presence at management team meetings.

In addition, in HaHP, some key aspects of the project were not covered by the external evaluation and struggled to get funded evaluation programmes in place. Examples of this were the cardiac rehabilitation programme and the development of a CHD disease register which was intended to provide the basis of a systematic approach to secondary prevention. HaHP submitted unsuccessful bids to the CSO and to the British Heart Foundation for funding for these evaluations. Subsequently, they developed an ambitious internal evaluation that is currently underway.

It is likely that the effectiveness of the cardiac rehabilitation programme and of the disease register in ensuring the systematic implementation of evidence-based interventions will be of

particular interest to the NHS throughout Scotland. In retrospect, the effective evaluation of this key element of HaHP should have been given a higher priority from the start and resourced adequately.

3.4 Relationship between internal and external evaluations

The experience of the three internal evaluations highlights the following 3 issues:

Evaluation and implementation role

At the outset, the focus of the DPs was defined as implementation/action and disseminating good practice; research and evaluation was not seen as part of these roles and there were expectations that the external evaluation team would do the research/evaluation. In particular, the requirements for internal monitoring and evaluation were unclear at the outset and lacked designated and clear strategic leadership.

The relationship between internal and external evaluation

There was a lack of clarity at the outset about the need for, and the nature and level of the internal evaluation function on the part of the DPs and the Scottish Executive. This meant that internal and external roles were not planned as complementary, integrated functions but kept quite separate. From HR's perspective, there had always been doubts about how the internal and external evaluation would fit together, and concerns about what would happen if there were a mismatch between the two. However, by commissioning the external evaluations first and independently from the projects, the internal evaluations were by default left doing everything else. HaHP felt that a clear understanding of what was required from the internal evaluation, and what resources were needed, should have been established at the outset. The lack of integration between internal and external evaluation roles also meant that there is no mechanism for bringing together and synthesising information from the internal and external evaluations. The HaHP team believed that it must be integrated much better than it has been in the past. Starting Well's view was that it was important for all the learning to be synthesised rather than focusing disproportionately on the results from the external evaluation.

Evaluation capacity and culture within implementing organisations

The whole process of uncertainty and lack of clarity and leadership for the internal evaluation function was compounded by the lack of capacity of some groups of staff to be 'critical practitioners' and to understand and build an internal evaluation role into the planning, development and review cycle. NHS secondary care was one exception, in that they had a strong culture of reflective practice and were developing information systems to support this. Variations in the evaluation culture across sectors were also noted by HR - the voluntary sector were seen to be very familiar with critical reflection and producing reports for funders; local authorities were seen to have an 'inspection culture' and produce committee reports when necessary; while the NHS were seen to have a more 'academic' approach.

4. Conclusions and recommendations

It is widely acknowledged (Macintyre et al., 2001; Wanless 2004) that evidence for effective interventions to improve health, and in particular to reduce health inequalities, is scarce. It is crucial that innovative interventions are accompanied by good quality evaluation (Macintyre, 2003). We start (4.1) by making some general recommendations for future commissioning and evaluation of complex community interventions such as HaHP, SW and HR, then (4.2) make some more specific recommendations for evaluation of phase 2 of the present demonstration projects, and finally (4.3) make some general points based on what we have gleaned from documents and interviews about the experience of the evaluation so far.

4.1 General recommendations

4.1.1 Commissioning projects

Alternative methods of commissioning should be considered, that place a greater weight on the existing evidence base, and take account of questions of 'evaluability'. For example, more resources should be devoted to the formal appraisal of options prior to commissioning. As well as informing project commissioning, this should also provide a clearer picture of expected impacts that can be used to inform the design of evaluations.

Future bids for DPs should be assessed by at least two independent expert reviewers who are not part of the commissioning group, and who can be asked specifically to comment on the strength of the evidence base.

A clear distinction should be drawn between phases of project development and implementation, and time allowed in the project's lifespan for the interventions to develop and settle down, before they are regarded as mature interventions amenable to a formal process and outcome evaluation. It is difficult however to let interventions mature when they are only funded on a short-term basis, in this case three years.

4.1.2 The evidence base

Allowing for the general weakness of the evidence base, the evidence for each of the three DPs was less convincing than presented in the project bids, and in one case was virtually ignored. Though some problems, such as transferability, were acknowledged, much better use should be made of existing evidence. We recommend:

- That the evidence base for the proposed interventions be made explicit at the commissioning stage;
- That evidence about the effectiveness of the proposed intervention be distinguished from background information about prevalence of risk factors, morbidity and mortality rates, etc;
- That the evidence base for the intervention be kept up to date during implementation, with particular attention given to any emerging evidence, both by the evaluators and by the projects;
- That consideration be given to the question of the transferability of the evidence base to a Scottish context, with particular attention to the extent to which evidence of intervention effects might be context dependent and therefore inapplicable – or not directly applicable - to the target population;
- That the differences between the design of the interventions represented in the evidence base and the design for the proposed intervention be clearly identified and

justified. The key is explicit clarity over what is evidence-based, what is new and what is an application to a new context.

4.1.3 Evaluation models

While there are advantages and disadvantages with both traditional and newer approaches we recommend:

- That the 'theory of change' be employed in the formative stages of future DPs where appropriate, not just in reactive mode but in shaping DPs and formulating a design that is 'evaluable'. This implies much earlier involvement of the evaluation team than was the case in the three current projects and a greater influence on their content and direction (e.g. recommending dropping particular components);
- That once objectives have been clarified and agreed, and the project is 'mature', a more traditional model of evaluation is appropriate, combining quantitative (e.g. quasi-experimental) methods to measure outcomes, continually informed by qualitative methods to focus on processes;
- That ways of retaining independence, while enabling researchers to get involved in project planning and development, should be considered, for example project steering groups on which non-grantholders are represented.

4.1.4 Commissioning evaluations

While the success or otherwise of the current evaluation studies cannot be adequately assessed until they are complete, the evaluators have clearly had to tackle a range of difficulties, some of which could and should have been anticipated at the commissioning stage. For future commissioning exercises we recommend that:

- The commissioning process should incorporate a strong element of independent peer review, as well as assessment by an internal tender panel;
- Planning for evaluation should begin at the same time as planning for the DP itself, so that evaluation considerations can be fed into the design and commissioning of the DP;
- Measures should be taken to ensure an adequate number of bidders. For complex or large scale evaluations, this may involve approaching potential bidders from throughout the UK as well as within Scotland, and bidders should be encouraged to form consortia to ensure adequate capacity, range of skills and expertise, etc;
- A clear distinction should be drawn between formative and outcome evaluation (see 4.1.3 above);
- Formal mechanisms for involving project stakeholders, including the evaluation such as involvement in project steering committees, should be considered;
- Project steering committees should also include independent scientific experts (i.e. non-grantholders);
- To keep evaluation tasks manageable, it is worth distinguishing questions about overall success and failure, which are difficult to answer precisely and involve a strong element of judgement, from questions about specific changes or impacts;
- Final reports should also be independently peer-reviewed, ideally by the same reviewers as the evaluation proposals.

4.1.5 Internal and external evaluation

We recommend that the respective roles of internal and external evaluation and monitoring should be clearly defined and agreed and a joint plan produced at the outset. This would help

avoid the gaps and overlaps and would help critical interdependencies (e.g. data sharing) to be identified, and resource requirements incorporated in project bids. The interactive nature of ToC (see above) should help to clarify roles and responsibilities in relation to evaluation. This will reduce the need of DPs themselves to engage in internal evaluation which may be of uneven quality and as a consequence of poor communication produce conflicting evidence. We suggest the following division of labour:

Internal role – monitoring
The focus of the internal evaluation function should be on:

- Setting up internal management information systems to enable the monitoring of project inputs (e.g. budgets & staff), outputs and participation/uptake/use;
- Assessing reach – surveying the project's target population to assess awareness, understanding, exposure and uptake/use of programme services/activities;
- Developmental research – research with users, non-users and local agencies/professionals to assess how the project is perceived and how marketing, communications and services/activities can be improved.

External role – evaluation
The focus of the external evaluation function should be on:

- Evaluating programme results, impacts and intermediate outcomes: what changes have resulted as a consequence of programme activities/services?
- Understanding how programme implementation has contributed to achieving the outcomes intended (process evaluation).

4.1.6 Capacity building

The Task Group's recommendations above suggest that there is a need to develop skills and capacity in a number of key areas if we are to realise improvements in the future. The main areas that need to be addressed are:

1. At national level, there is a need to develop a more sophisticated approach to commissioning policy-related initiatives from local partnerships that would enable a developmental stage to be built in. During this stage, the most promising initial proposals can be worked up into full detailed bids in collaboration with funders, local partners, relevant academic experts and evaluators. At this early stage of intervention development, the evidence base for the initiative could be considered and an evidence 'template' drawn up, together with the programme theory (or ToC) and plans for evaluation.
2. At national and local level, there is also a need to strengthen skills in developing health improvement plans that are informed by the existing evidence, and which are evaluable and contribute to a learning agenda. This may require a combination of written guidance, training and ongoing advice and support. The development of a new inter-sectoral evaluation service for Scotland may be able to contribute towards this development agenda.

If the commissioning process and the evidence-basis and evaluability of projects were thus improved, this may also help to address the evaluation capacity problems experienced in relation to the academic research community. The reluctance of academic researchers to engage in the evaluation of complex community initiatives is largely related to their lack of

involvement at an early stage in the intervention development process, the poor evidence basis of the initiatives, and their low evaluability.

4.2 Recommendations for Phase 2

The external evaluations of the 3 projects should not continue in their current format. Specific recommendations for the three projects are as follows:

4.2.1 Healthy Respect

There could be some justification for continuing evaluation of the schools element of HR. While the evidence from the SHARE trial to date does not suggest an effect on behaviour, further data on longer term outcomes (termination of pregnancy) is expected soon. In Phase 2, HR plans to introduce SHARE as part of its overall programme of supporting education and services in schools across a whole authority. This could be rolled out in 2 stages over 2 school years, with half the schools implementing SHARE in year 1, the other half in year 2, thus enabling the latter to serve as control schools in year 1. Because all schools in the local authority will be involved, this would avoid the problems of selection experienced in the evaluation to date, especially if schools were randomised to early and late intervention.

4.2.2 Starting Well

There is no justification for continuing the external evaluation of either SW on anything like the present scale, a view shared by the external evaluation team. The intervention is still changing so rapidly that it is impossible to know what is being evaluated. There may be a case for continuing to follow up the SW cohorts, but this can only be properly assessed once response to the 18 month follow-up is known.

4.2.3 Have a Heart Paisley

Likewise, continued large-scale external evaluation of HaHP is unjustified, given problems with non-response and the consequent risk of bias in the baseline surveys. While there seems little point continuing with the original attempt to measure population level outcomes, a focus on a small number of agreed priority areas, where sustainable approaches have or are being developed may be useful.

4.2.4 Internal evaluation and monitoring

This should continue, as part of good project management, but Phase 2 bids should include costed proposals, specifying questions, data sources and deliverables.

4.3 General conclusions

Although the Task Group's remit was evaluation, some more general issues emerged in our discussions with the evaluation and DP teams which are worth briefly mentioning. We regard these as potentially important for future projects of this kind, both to ensure that scarce skilled staff are used appropriately and to provide good value for the substantial sums of public funds invested.

The concept of a 'demonstration project' is obscure and problematic and despite attempts to clarify definitions, it still remains a fundamental issue and is likely to be a matter of continuing debate.

The commissioning process should be organised in such a way that the evidence base for the proposals (and any significant gaps), is clear and explicit before contracts or grants are

awarded. Where evidence is weak, a formative stage should be explicitly built into the project's timetable, and evaluation commissioned accordingly.

The timescales for the DPs were unrealistic and uncoordinated, leading to real problems for both the project teams and the evaluators and ourselves. There seems to have been very little consideration until late in the day of an exit strategy, incorporating time to review and assimilate the results of the evaluations. It is vital that this is considered early and more thoroughly in Phase 2. Likewise, a realistic, fully coordinated timetable for the implementation and funding of any future DPs and their evaluations is essential, to ensure that the appropriate evidence is available for informing decisions about the projects and wider policy within Scotland.

This report is not conclusive, since the evaluations themselves are still in progress, but we hope it will trigger further constructive debate. Our final recommendation is that a series of workshops are held, for projects, external teams, learning networks, Co-ordinating Group and Evaluation Task Group members, and HISD colleagues to discuss further the issues we have identified.

REFERENCES

Barnard, K. E., Magyary, D., Sumner G. et al. (1988), Preventing parenting alterations for women with low social support, *Psychiatry*, 51, 248-253.

Blamey, A. (2001) *Have a Heart Paisley: Report on the Theories of Change Development*. Health Promotion Policy Unity, University of Glasgow, August 2001

Blamey, A. (2002) *The evidence base for community-based CHD interventions.* Health Promotion Policy Unit, University of Glasgow

Blamey, A. (2003) *Delivering on Expectation? Have A Heart Paisley's Progress against initial plans. Year 2 Theory of Change evaluation. Interim Report.* Health Promotion Policy Unity, University of Glasgow, July 2003.

Bradley, R., & Caldwell, B. (1979), Home observation for the measurement of the environment: a revision of the pre-school scale, *Amer. J. Ment. Defic.,* 84, 235-244.

Cabinet Office (2003) *The Magenta Book: Guidance Notes for Policy Evaluation and Analysis*. Government Chief Social Researcher's Office, Prime Minister's Strategy Unit, Cabinet Office, July 2003

Carleton RA, Lasater TM, Assaf AR *et al.* The Pawtucket Heart Health Program: community changes in risk factors and projected disease risk. *Am J Public Health* 1995; **85:**777-85.

Chelimsky, E. (2001) Introduction. *Twenty-first Century Evaluation*. Sage Publications.

Clinical Evidence; **Issue 9**: June 2003 p166-202. BMJ Publishing Group

Ebrahim S, Davey Smith G, Exporting Failure? Coronary Heart Disease and stroke in developing countries. *Int J Epidemiol* 2001; **30:** 201-205

Ebrahim S, Davey Smith G. Systematic review of randomised controlled trials of multiple risk factor interventions for preventing coronary heart disease. *BMJ* 1997; **314:** 1666-1674

European Commission (2003) *Evaluating EU Activities: A Practical Guide for the Commission Services*. EC DG Budget, Brussels. November 2003. http://europa.eu.int/comm/budget/evaluation/

Farquhar JW, Fortmann S, Flora J *et al.* Effects of community wide education on cardiovascular disease risk factors. The Stanford Five-City Project. *JAMA* 1990;**264:**359-65

Fullbright-Anderson, A., Kubisch, A. C. & Connell, J. P. (1998). New approaches to evaluating community initiatives: theory, measurement and analysis. Washington DC: The Aspen Institute.

Gutelius, M. F., Kirsch, A. D., Macdonald, S., Riddick Brooks, M. & McErlean, T. (1977), Controlled study of child health supervision: behavioral results, *Pediatrics*, 60(30, 294-304.

Hanlon, P., Morrison, C., Ford, I., Judge, K., Reid, M., Gilhooly, M. (2000), A bid to evaluate the 'Have a Heart Paisley' National Demonstration Project

Hardcastle JD, Chamberlain JO, Robinson MHE, Moss SM, Amar SS, Balfour TW et al. Randomised controlled trial of faecal occult blood screening for colorectal cancer.Lancet 1996, 348; 1472-1477

HEBS (1999) *Research for a Healthier Scotland: The research strategy of the Health Education Board for Scotland 1999–2004*. HEBS, Edinburgh.
http://www.hebs.com/research/

Hodnett, E. D. & Roberts, I. (1999), Home-based social support for socially disadvantaged mothers, The Cochrane Database of Systematic Reviews, 4(4).

Johnson, Z., Howell, F. & Molloy, B. (1993), Community mothers' programme: randomised controlled trial of non-professional intervention for parenting, *Brit Med J*., 306, 1449-1452.

Judge, K. & Bauld, L. (2001), Strong theory, flexible methods: evaluating complex community-based initiatives, *Critical Public Health*, 11(1), 19-38.

Kendrick, D., Elkan, R., Hewitt, M., Dewey, M., Blair, M., Robinson, J. et al., (2000), Does home visiting improve parenting and the quality of the home environment? A systematic review and meta analysis. *Arch. Dis. Child*., 82, 443-451.

Kirby, D. (1999), sexuality and sex education at home and school, *Adolescent Medicine*, 10(2), 195-209.

Kirby D., Short, L., Collins, J., Rugg, D., Kolbe, L., Howard, M. et al (1994), School-based programs to reduce sexual risk behaviors: a review of effectiveness, *Public Health Reports*, 109(3), 339-360.

Kronborg O, Fenger C, Olsen J, Jorgensen OD, Sondergaard O. Randomised study of screening for colorectal cancer with faecal occult blood test. Lancet 1996, 348; 1476- 1471.

Lealman, G., Haigh, D. Phillips, J. et al. (1983), Prediction and prevention of child abuse – an empty hope? *Lancet*. (1), 1423-1424.

Luepker R, Murray D, Jacobs DR *et al.* Community Education for cardiovascular disease prevention: risk factor changes in the Minnesota Heart Health Program. *Am J Public Health* 1994; **84:**1383-93.

McAlister FA, Lawson FM, Teo KK, Armstrong PW. Randomised trials of secondary prevention programmes in coronary heart disease: a systematic review. *BMJ* 2001; **323:**957-962

Macintyre, S., Chalmers, I., Horton, R. & Smith, R. (2001), Using evidence to inform health policy: case study, *Brit Med J.,* 322, 222-225.

Macintyre, S. (2003), Evidence based policy making: impact on health inequalities still needs to be assessed, Brit med J., 326, 5-6 (editorial).

Mackenzie M and Blamey A (2004) (in press) The practice of a theory based approach: Lessons from the application of a theory of change, *Evaluation*

Mackenzie, M (2002) *Theory of Change and implications for implementation and evaluation,* Glasgow: HPPU

Mackenzie, M. (2003) *Implementation of an intensive home visiting model within the Starting Well Demonstration Project. Interim Report*. Health Promotion Policy Unity, University of Glasgow, April 2003.

MacMillan, H. L., MacMillan, J. H., Offord, D. R., Griffith, L. & MacMillan, A. (1994), Primary prevention of child ophysical abuse and neglect: a critical review. Part 1*, J. Child. Psychol. Psychiat*, 35(5), 835-856.

Mandel JS, Bond JH, Church JR, Snover DC, Bradley GM, Schuman LM et al. Reducing mortality from colorectal cancer by screening for faecal occult blood. N Engl J Med 1993; 328; 1365-1371

Meelianby, A. R., Phelps, F. A., Crichton, N. J. & Tripp, J. H. (1995), School sex education: an experimental programme with educational and medical benefits, *Brit Med J.,* 311, 414-417.

Minnesota Heart Health Program. *Am J Public Health* 1994; **84:**1383-93.

NHS Centre for Reviews and Dissemination (1997), Preventing and reducing the adverse effects of unintended teenage pregnancies, Effective Health Care, 3(1), 1-12.

Olds, D. L., Eckenrode, J., Henderson, C. R., Kitzman, H., Powers, J., Cole, R. et al. (1997), Long-term effects of home visitation on maternal life course and child abuse and neglect: fifteen-year follow-up of a randomized trial, *J. Amer. Med. Assoc*., 278(8), 637-643.

Olds, D. L., Henderson, C. R., Chamberlain, R & Tatelbaum, R. (1986). Preventing child abuse and neglect: a randomized trial of nurse home visitation, *Pediatrics*, 78(1), 65-78.

Olds, D. L., Henderson, C.R. & Kitzman, H. (1986), Does prenatal and infancy nurse home visitation have enduring effects on qualities of parental caregiving and child health at 25-50 months of life? *Paediatrics*, 93, 89-98.

Puska P. Evaluation. In Puska P, Tuomilheto J, Nissinen A, and Vartiainen E (1995). *The North Karelia Project. 20 Year Results and Experiences.* Helsinki, National Public Health Institute

Puska P. Do we learn our lessons from the population-based interventions? *J Epid Community Health* 2000; **54:**562-563.

Puska P, Nissinen A, Tuomilehto J *et al.* The community based strategy to prevent coronary heart disease. Conclusions from the ten years of the North Karelia project. *Ann Rev Public Health* 1985;**6:**147-93.

Puska P, Tuomilheto J, Nissinen A, and Vartiainen E (1995). *The North Karelia Project. 20 Year Results and Experiences.* Helsinki, National Public Health Institute

Sanders, M. R. (1999), Triple P- Positive Parenting Programme: Towards an empirically validated multilevel parenting and family support strategy for the prevention of behavior and emotional problems in children, *Clin,. Child. Fam. Psychol. Rev*., 2(2), 71-90

Scottish Executive (2003), *Improving Health in Scotland – the Challenge*

Scottish Office (1999), *Towards a Healthier Scotland*

Shute, J. & Judge, K. (in press), Evaluating 'Starting Well', the Scottish Demonstration Project for Child Health: outcomes at 6 months.

Stern, E. (2004) *Action Research and Evaluation Practice*. Scottish Evaluation Network Workshop, Edinburgh, 30 Jan 2004.

Todd J.Currie, G.& Smith,R.(1999), Health Behaviours of Scottish schoolchildren: Technical report 2 -Sexual health in the 1990s. Research unit in Health and behavioural change, Edinburgh: University of Edinburgh.

Tucker,J.,van Teijilingen,E.,Shucksmith.J & Penney,G.(2000), Evaluation for the Lothian Healthy Respect demonstration Project (revised proposal), August, 2000

Tudor-Smith C, Nutbeam D, Moore L *et al.* Effects of the Heartbeat Wales programme over five years on behavioural risk for cardiovascular disease: quasi-experimental comparison of results from Wales and a matched reference area. *BMJ* 1998; **316:** 818-822.

Vartiainen E, Puska P, Pekkanen J, *et al.* Changes in risk factors explain the changes in mortality from ischaemic heart disease in Finland. *Br Med J* 1994; **309:** 23-27.

Wanless, D (2004). *Securing Good Health for the Whole Population*. London, HM Treasury (http://www.hm-treasury.gov.uk/consultations_and_legislation/wanless/consult_wanless04_final.cfm).

Weiss, C.H. (1995) Nothing as practical as a good theory. In: Connell et al (eds) *New Approaches to Evaluating Community Initiatives: Concepts. Methods and Contexts.* Washington DC, Aspen Institute.

Wellings, K., Wadsworth, J., Johnson, A. M., Field, J., Whitaker, L. & Field, B. (1995), Provision of sex education and early sexual experience: the relation examined, *Brit Med. J.,* 311, 417-421.

West P. & Sweeting, H. (1997), Lost souls and rebels: a challenge to the assumption that low self-esteem and unhealthy lifestyles are related, *Health Education*, 5, 161-167.

Wight, D, Raab, G., Henderson, Abraham, C. Buston, K., Hart, G et al. (2002), The limits of teacher-delivered sex education: interim behavioural outcomes from a randomised trial, *Brit Med J*., 324, 1430-1430.

Wimbush, E. and Watson, J. (2000) An evaluation framework for health promotion: theory, quality and effectiveness. *Evaluation*. Vol. 6 (3): 301-321.

REFERENCES FOR INTERVIEWS

Interviews were held by the Evaluation Task Group as follows:

Have Heart Paisley/Starting Well Evaluation teams – 9 October 2003

Healthy Respect Evaluation team – 4 November 2003

Healthy Respect team – 6 November 2003

Starting Well team – 7 November 2003

Have a Heart Paisley team – 2 December 2003

National Learning Networks – 16 December 2003

ANNEX A

THE COLORECTAL CANCER SCREENING PILOT –the Fourth Health Demonstration Project

1. Background

Colorectal cancer is the second commonest cause of cancer death in the UK, with over 3000 new cases and 1700 deaths each year in Scotland.

Three population based randomised trials [1,2,3] have shown conclusively that screening using faecal occult blood testing (FOBt) can reduce death rates from colorectal cancer by about 20%. However, such trials are tightly controlled studies carried out by highly motivated individuals and there was some doubt as to whether similar results could be obtained in a routine setting within the NHS.

2. Developing the Pilot

The National Screening Committee (NSC) held two workshops in 1997 and 1998 to canvas a wide range of expert opinion and critically review the available evidence. It was recommended that the feasibility of a national screening programme should be tested before committing the UK Health Departments to such a programme. To do this, a demonstration project (DP) based on two populations of about one million each, in different parts of the country was proposed. An outline specification for the project was agreed – including the target age range (all 50-69 year olds) and the technique to be used (biennial FOB testing with colonoscopy as the follow up investigation for those testing positive).

When this national specification had been approved, all UK Health Authorities were invited to submit expressions of interest in hosting a pilot site. 5 proposals were shortlisted for further consideration by an assessment panel appointed by NSC. Members of this panel visited these sites and, in early 1999, the final selection was made after presentations and interviews.

The sites chosen for the pilot were North East Scotland (Grampian, Tayside and Fife) and Coventry and Warwick in England. An Executive group was set up to oversee the operational development and running of the pilot, comprising the National Screening Coordinators for Scotland and England, their central project managers, the two clinical directors and the two pilot managers. A UK steering group was also set up to formulate general policy.

Working groups were set up to develop the specification for the computer system, to select the FOB test, to develop data sets, protocols, quality assurance systems and information materials.

The first round of screening began in March 2000. FOBt kits were sent by post to the target population, with instructions to complete the test and to send it back to the screening centres in Dundee and Coventry for analysis.

3. Evaluation Model and Methods

3.1 External evaluation

Independent evaluation was an essential part of the project and a single evaluation covering both the Scottish and English arms of the pilot was agreed. This was put out to tender by the R&D Directorate of the Department of Health in London. On the basis of competitive bids, the evaluation was awarded to a group from the Universities of Edinburgh, Warwick and Essex.

The evaluation covered many aspects of the pilot: performance against benchmarks, the impact on primary and secondary care services, uptake and the reasons for non – participation, physical and psychological morbidity associated with screening, and an economic evaluation.

The evaluation was complex and required a multi-disciplinary approach, including epidemiology, health services management, psychology, health economics, primary and secondary care. The evaluation team were located at three sites, and needed to interact with the two pilots and with individuals in the Scottish Executive and the Department of Health.

Developing a carefully managed relationship with the pilot sites was essential. The evaluation had to be independent and unbiased, yet good communication with the pilot sites was essential – e.g. to ensure agreement on definitions and on quality standards, and to coordinate the evaluation surveys with the screening programme. For these reasons, the evaluators developed a 'Terms of Engagement' document which aimed to make explicit the competing imperatives of independence and integration. This was found useful by all concerned.

In February 2003, the Evaluation Group produced an independent report for the Department of Health R&D Directorate on the outcomes of the pilot, with the aim of informing a decision on a national programme.

3.2 Internal monitoring/evaluation

Initially, insufficient priority and resources were given to developing the internal monitoring capacity. There was an implicit assumption that the external evaluation would address all the issues. When this was identified as a priority, there were some problems with timely access to the data and ensuring the IT system could provide the necessary information.

4. Results of the Evaluation

A decision on whether to continue the pilot for a second round was needed during 2002, well before the results of the evaluation were available. A second round of screening began in Scotland in December 2002.

The final report of the evaluation group was submitted to the Department of Health in February 2003. It showed that the UK pilot had demonstrated that key parameters of test and programme performance observed in randomised studies of FOBT screening can be repeated in population based pilot studies. The cost effectiveness also compared well with other forms of cancer screening.

The independent external evaluation has not been continued to cover the second round of screening. The internal monitoring capacity has been enhanced to address this.

In July 2003, the Scottish Executive gave a commitment in principle to roll out a national colorectal screening programme, while recognising that planning for its introduction could take up to 5 years.

5. Key Issues

- As a first stage in developing the demonstration project, national workshops were held to canvas expert opinion and critically review the literature. A detailed specification setting out core elements of the project, with supporting documentation, was then agreed nationally and subsequently put out to tender. Decisions on the successful proposals were made by an expert group, on the basis of detailed plans for implementation. This model differed significantly from that used for the other three demonstration projects.

- There was a mismatch in timescales resulting in decisions about the future of the project being needed before the results of the independent evaluation were available. Similar difficulties were faced by the other three demonstration projects.

- The external evaluation team recognised the need for a balance between independence from and engagement with the project team. This issue was shared by the other demonstration projects and the right balance will vary depending on the circumstances. The need for explicit 'rules of engagement' is important and recognised by all participants.

- The importance of adequately resourced internal monitoring systems was not initially identified as a priority, but was subsequently addressed successfully. This was also an issue for the other demonstration projects.

- The external evaluation was concluded with the final report of the team in February 2003. Evaluation of subsequent screening rounds of the project will be mainly through internal monitoring systems. Specific focussed research proposals are also being developed, seeking external funding through the usual channels.