How can the health equity impact of universal policies be evaluated?

World Health Organization
REGIONAL OFFICE FOR Europe

How can the health equity impact of universal policies be evaluated?

Insights into approaches and next steps

Edited by:
Beth Milton
May Moonan
David Taylor-Robinson
Margaret Whitehead

WHO Collaborating Centre for Policy Research on Social Determinants of Health,
University of Liverpool.

World Health Organization
REGIONAL OFFICE FOR Europe

UNIVERSITY OF LIVERPOOL

How can the health equity impact of universal policies be evaluated?

Insights into approaches and next steps

Synthesis of discussions from an Expert Group Meeting

A joint publication between the WHO European Office for Investment for Health and Development, Venice, Italy
and the
Liverpool WHO Collaborating Centre for Policy Research on Social Determinants of Health

Held at the University of Liverpool, 2–4 November 2010

Edited by
Beth Milton, May Moonan, David Taylor-Robinson, Margaret Whitehead

ABSTRACT

Taking population level action on the wider social determinants of health in efforts to reduce health inequities is an international public health imperative. However, an important barrier to action is the perceived lack of evidence about what works to reduce health inequities. This is particularly evident in relation to universal welfare policies, which can have profound effects on health inequities, both positive and negative in nature. Because universal policies are usually applied to whole populations, and are often complex in nature with long causal chains, this precludes a true experimental design, and other approaches to evaluation are required. This report presents arguments and case studies from an expert group meeting convened to clarify the importance and challenges of evaluating universal policies, and to outline potential approaches to assessing the impact of universal policies on health inequities. The report also identifies key research and policy questions that need evaluating as a matter of priority, and sets the agenda for partnership working to develop these methods further.

Keywords

SOCIOECONOMIC FACTORS
HEALTH SERVICES ACCESSIBILITY
HEALTH POLICY
UNIVERSAL COVERAGE
PROGRAMME EVALUATION
EUROPE

Address requests about publications of the WHO Regional Office for Europe to:
Publications
WHO Regional Office for Europe
Scherfigsvej 8
DK-2100 Copenhagen Ø, Denmark
Alternatively, complete an online request form for documentation, health information, or for permission to quote or translate, on the Regional Office web site (http://www.euro.who.int/pubrequest).

Contents

Acknowledgements

We wish to thank Chris Brown for her contribution to the content and design of the event; Yelan Guo for her preparatory work for this meeting and for her role as rapporteur during the meeting. Many thanks also to Antonella Biasiotto and Yvonne Chantler for their administrative support in the preparation for and running of the meeting.

The views expressed and discussed in this report are those of the participants and are not necessarily representative of views of the governments, ministries or other stakeholders in the countries concerned.

List of abbreviations

CMO	Context-Mechanism-Outcome
EU	European Union
GDP	Gross domestic product
ILO	International Labour Organization
OECD	Organisation for Economic Co-operation and Development
OPEC	Organization of the Petroleum Exporting Countries
NDC	New Deal for Communitites
NHS	National Health Service
RCT	Randomized control(led) trial
WHO	World Health Organization

Participants

Bo Burström
Department of Public Health Sciences
Karolinska Institute
Sweden

Tatjana Buzeti
Centre for Health and Development
Murska Sobota
Slovenia

Mark Exworthy
School of Management
Royal Holloway, University of London
United Kingdom

Maciek Godycki-Cwirko
Department of Family and Community
Medicine
Medical University of Lodz
Poland

Peter Goldblatt
Department of Epidemiology and Public Health
University College London
United Kingdom

Beth Jackson
Strategic Initiatives and Innovations Directorate
Public Health Agency of Canada
Canada

Ken Judge
School for Health
University of Bath
United Kingdom

Mike Kelly
Centre for Public Health Excellence
National Institute for Health and Clinical Excellence
United Kingdom

Rene Loewenson
Training and Research Support Centre
(TARSC)
Zimbabwe

Olle Lundberg
Centre for Health Equity Studies (CHESS)
University of Stockholm
Sweden

Mark Petticrew
Faculty of Public Health & Policy
London School of Hygiene & Tropical Medicine
United Kingdom

Jennie Popay
Division of Health Research
Lancaster University
United Kingdom

Tone Poulsson Torgersen
Norwegian Directorate for Health
Norway

Barbara Rohregger
WHO European Office for Investment for Health and Development
Italy

Alan Shiell
Population Health Intervention Research Centre
University of Calgary
Canada

David Stuckler
Department of Sociology
University of Oxford
United Kingdom

Marc Suhrcke
School of Medicine, Health Policy and Practice
University of East Anglia
United Kingdom

Carol Tannahill
Glasgow Centre for Population Health
United Kingdom

WHO Regional Office for Europe
Richard Alderslade
Division of Country Health Systems
Denmark

WHO Headquarters
Nicole Britt Valentine
Ethics, Equity, Trade and Human Rights
World Health Organization
Switzerland

Margaret Whitehead *(convener of meeting)*
WHO Collaborating Centre for Policy Research on Social Determinants of Health
Institute of Psychology, Health and Society
University of Liverpool
United Kingdom

Rapporteurs
May Moonan
Institute of Psychology, Health and Society
University of Liverpool
United Kingdom

David Taylor-Robinson
Institute of Psychology, Health and Society
University of Liverpool
United Kingdom

Yelan Guo
WHO European Office for Investment for Health and Development
Italy

Section 1: Introduction

1.1 Background

One of the burning issues in public health internationally is the recognition of the need to take action on the wider social determinants of health in efforts to reduce health inequities. As a consequence, the World Health Organization (WHO) global Commission on Social Determinants of Health called for health equity assessments of all policy reforms across all government sectors (CSDH, 2008). This is also reinforced in the European Union's (EU) 2009 "Council conclusion on solidarity in health: reducing health inequalities in the EU" (European Commission, 2009).

At the same time, it is often lamented that we lack evidence of what works to reduce health inequities – a conclusion based on reviews from controlled experiments with specific interventions (see, for example, House of Commons Health Committee, 2009). That pessimistic conclusion can be challenged if broader types of evidence are employed. There is no doubt, however, that there is a mismatch between what *should* be evaluated and what *is* being evaluated, as well as in the concepts and methods of evaluation employed: a micro-approach is predominantly adopted, when a macro-approach is required (Smith & Petticrew, 2010; Shiell, 2010; Whitehead, 2010).

The evaluation dilemma is most evident in relation to universal welfare policies, which, by definition, cover the whole of the population and therefore cannot be evaluated by the use of an unexposed control group within that population. Yet, such policies are potentially of great importance for reducing health inequities and promoting the well-being of citizens across the social gradient. Methods need to be developed and refined for assessing the impact of such universal policies on public health and health inequities.

The need to find ways to evaluate the health and inequities impact of universal policies has become more acute with the current economic crisis. The public sector in general is seen as an easy target for drastic cuts in many European countries, with universal welfare policies being particularly vulnerable, abandoned in favour of targeted services for the poor alone.

1.2 Purpose of the Expert Group Meeting in Liverpool

It is against this background that a meeting was proposed of a small expert group of people who have been tussling with the challenges of tracking the impact of wider social policies, together with stakeholders, who are involved in steering national strategies. The purpose was to explore the importance and feasibility of

evaluating universal policies and their impact on social determinants of health and the reduction of health inequities (for the full meeting agenda, see Appendix 1).

The meeting was co-organized by the WHO European Office for Investment for Health and Development (Venice, Italy) and the newly re-designated WHO Collaborating Centre for Policy Research on Social Determinants of Health (United Kingdom), and was hosted in Liverpool from 2 to 4 November 2010.

Specific objectives

- Clarify the importance of evaluating universal policies and the challenges.
- Present and review current methods for and approaches to assessing the impact of such policies on social determinants of health and/or health inequities, and discuss refinements.
- Identify "candidate" policies that could be assessed, including natural policy experiments and live policies in Europe as well as internationally.
- Explore interest in establishing a partnership between countries and institutions to evaluate universal policies, share findings and develop related products in order to support further action in this area.

Expected outputs

- Publication on approaches to evaluating population-wide policies for their health equity impact.
- A "shopping list" of research questions and policies that need evaluating as priorities.
- A plan for taking forward a partnership of researchers and policy advisers, with WHO, to inform national and international strategy on the social determinants of health and the reduction of inequities in health.

1.3 Structure of the report

The structure of the report follows the four objectives listed in subsection 1.2, synthesizing the discussions throughout the 3-day meeting under five headings that reflect the research questions implied in the objectives. [1]

1. Why is this issue important now?
2. What are the main challenges for evaluation?
3. What methodological approaches are promising/have been tried on live policies? How could these methods be refined/developed further in order to fill the gaps?
4. What are the burning research questions and policies that need evaluating as a matter of priority? Can some natural policy experiments be identified to address these questions?
5. How can we establish a longer-term partnership of researchers, policy advisers and WHO in order to advance the evaluation agenda?

Note: Both research and policy perspectives on each of these issues were considered to be essential and as such were incorporated into each session of the meeting.

[1] In sections 2–6, statements by participants are identifiable by indented italic font, with the relevant person's name attributed.

Section 2: Why is it important to evaluate "universal policies" for their impact on health inequities?

We use the term "universal policies" here to mean those that are applied to the whole population, but – crucially, in terms of our concern with health inequities – they may have differential impact or consequences. Some social policy commentators reserve the term "universal" for a subset of population-wide services to which everyone has access, which are free or generally affordable at the point of delivery, and which are underpinned by social cohesion principles. Our discussions took in the wider definition, encompassing not just *services* but also *policies*, such as the tax system, and considering those that may be regressive as well as those based on social cohesion principles. There are several reasons why a renewed drive to evaluate universal policies should be fostered at this point in time, as discussed in the subsections that follow.

2.1 Universal policies are potentially important determinants of health

First, universal policies can be considered to be determinants of health in themselves, as well as being seen as interventions that aim to ameliorate the consequences of the wider determinants of health.

> *We talked about universal policies in two quite distinct ways. First, as a determinant of health, and therefore as part of the evidence base, which helps us to understand the relationship between the wider determinants of health and various health outcomes. Second, we talked about universal policies as interventions to ameliorate the worse consequences of the wider determinants of health, and of course they are both of those things.* (Mike Kelly)

These policies have the capacity to exert a profound influence on health and health inequities, both positive and negative in nature; precisely because they affect many people, it is possible to detect a significant impact at the population level.

2.2 Countering the "inverse evidence law"

Second, such policies suffer from the "***inverse evidence law***" (Nutbeam, 2003), which means that there is the least evidence available to inform the implementation of those policies that have the greatest potential to influence population health.

> *The Global Commission made a special recommendation that equity impact assessment should be carried out for all policy-making and market regulation, nationally and internationally. But, the problem is what Don Nutbeam has called the 'inverse evidence law', and that is the availability of evidence tends to vary inversely with the potential impact of the intervention … So, why is there this inverse evidence law in relation to action on social determinants of health? Well, first the technical difficulty is huge as you go upstream, compared with the task of evaluating a small project*. (Margaret Whitehead)

Another reason for the inverse evidence law in relation to action on health inequities is the tendency towards "**lifestyle drift**".

> *Lifestyle drift: it's when you have policy initiatives aimed at tackling health inequities that start off with a broad social determinants or upstream approach, and then drift downstream to focus largely on individual lifestyle factors. This is often coupled with the drift away from recognition of the need to take action on the social gradient to a narrow focus on the most disadvantaged. We have seen this in the United Kingdom, in some of the case studies collected by WHO around Europe, and globally.* (Margaret Whitehead)

Lifestyle drift reflects the ways in which policy-makers approach complex social problems. The political emphasis on short-term outcomes and indicators facilitates lifestyle drift. Politicians are also prone to identifying behavioural interventions as a perceived solution to tackling the poor health of "high-risk" individuals.

2.3 The need for evidence to support efficient and effective use of scarce resources

Third, evidence is needed from evaluations of universal polices to inform the debate on the most efficient and effective use of scarce resources. The global financial crisis has intensified the debate regarding whether to provide publicly funded services for everyone or targeted services only for the poor. Assumptions relating to costs and benefits of different strategies are bandied about in support of one or other position, without a proper examination of the underpinning evidence base.

> *The trouble in a European context is that we start thinking 'target', we start thinking conditionalities, we start thinking pro-poor as opposed to universalism. And certainly in England now, there is a very real debate about 'If we are going to be pro-poor, then we don't have universal services'… which is why the evaluation agenda is so crucial, because these two policy arguments are being framed increasingly as being in competition with one another.* (Jennie Popay).

> *In a country like Poland, many people including public health professionals are not prepared for a general discussion, based – for example – on the rainbow*

model [Dahlgren & Whitehead, 1993]. And when you are talking about universal policies, universal approaches and universal methodologies, people just doubt it is even possible, so they don't undertake the discussion of it. (Maciek Godycki-Cwirko).

2.4 Piecemeal learning at the local level

Fourth, it is particularly challenging to implement population-wide policies at the local level. When local coalitions are built, resources tend to be dissipated among a large number of local partners, and aims to achieve a population impact risk becoming diluted into small, 'downstream' initiatives. An inability to develop convincing logic models, and poor design and implementation of interventions contribute to these failures at the local level.

When you build these local coalitions, you start off with a determination to use the resources you have got – to make a real difference. And in doing that, you need to build a coalition. But when you build a coalition of many different interests, what happens is you dissipate the resource across all the partners. Take one example from the health action zone in Tyne and Wear – a conurbation of 1.5 million. One of their objectives was a dramatic transformation and reduction of inequities in teenage pregnancy rates. But when you look at what they actually did, they had one refuge in Sunderland for a handful of young women. Another example is 'Have a Heart Paisley' – a coronary heart disease health prevention demonstration project in Scotland. This had very laudable objectives of promoting physical activity amongst people post-retirement, and wanted to make a population-level impact. But when you look at what they were actually able to do, they had one line dancing class for 30 people on a Wednesday afternoon that ran for six weeks. This was probably very good for the 30 participants, but it wasn't going to make a population-level impact. So, we have not suffered principally from a lack of evaluation resources or expertise. The main failure is related to the poor quality of design and implementation. Objectives have either been hopelessly over-ambitious at the strategic level, or too opaque at the operational level, or both … Resources, have been too easily dissipated and – as a result – learning has been modest and piecemeal. (Ken Judge)

2.5 The current erosion of universal policies

Many governments are currently reviewing their universal policies and services as part of cost-cutting programmes in response to the global economic recession. It is therefore urgent to assess the impact of universal policies on the social determinants of health. It is also important to demonstrate the value of universal policies more broadly.

> *Globally the current economic crisis will have an adverse influence on these [universal] policies and so the need is as urgent as it's ever been to be able to track, and to make visible, what the policy impacts are* (Margaret Whitehead).

> *We are entering treacherous waters globally: there are indiscriminate cuts in public services, and these cuts are likely to have differential consequences, hitting the worst-off in society the hardest* (Margaret Whitehead).

2.6 Better evidence for policy-making

A larger and stronger body of evaluation evidence for the effectiveness of universal policies may enable politicians to develop and clarify their political intentions. A better developed evidence base may demonstrate to them how to link political ideology with their desired social and health outcomes.

Section 3: What are the evaluation challenges and barriers in relation to such policies?

3.1 The issue of controls/comparison groups

The defining characteristic of a universal policy – that it is applied across the whole population – presents one of the biggest challenges to evaluation of effectiveness because it almost always precludes a true experimental design in which a comparison is made with an unexposed control group. As the whole population is "exposed", then great ingenuity is required to devise appropriate, robust approaches to assessing impact. Approaches developed so far have included comparisons with populations in other countries, or in other regions, or at different times in the same population before and after exposure, taking advantage where appropriate of "natural policy experiments". A range of techniques is then needed to take account of differences in context and composition of the populations being compared, to ensure the proper interpretation of results.

3.2 Time-lags and long causal chains – temporal issues for evaluators

Donald Campbell's paper (1969) on the potential for natural experiments to inform policy decisions argued that it is important to learn lessons about the impact of every policy reform. The challenges that he outlined to the evaluation of natural experiments include:

- "history/maturation" bias (confounding with other interventions);
- "instrumentation" bias (measurement/ascertainment biases);
- "regression artefacts" (regression to the mean, since interventions are often brought into play whenever a problem is at its height and then it naturally declines);
- "experimental mortality" (attrition);
- biases common to any observational study;
- complex causal webs involved;
- complex interventions that are not discrete "on–off" interventions that can be easily described, but are phased, ill-defined, messy, and involve adaptive feedback loops;
- multiplicity of significant outcomes, of which health is just one (there is a risk of undervaluing an intervention by just focusing on health outcomes).

It is therefore clearly important that researchers are aware of and make every effort to minimize these different kinds of bias in an attempt to ensure that the

evidence produced during the evaluation of natural experiments is not misleading. It is possible that econometric techniques to deal with potential selection bias in natural experiments may offer a means of tackling some of these issues.

The long time-lags that may occur between a universal policy being implemented and health or social determinants-related outcomes being observed are a key challenge for those who seek to demonstrate the effects of universal policies. It is difficult to conduct a prospective evaluation of an intervention when it may take a generation to see an impact on health-related outcomes. The long time-lags between interventions and health or social determinants impacts may also act as disincentives for politicians to invest in the implementation of universal policies, or in their evaluation, because of the need to see short-term changes within tight political time frames.

> *The complexity of the causal changes – it's difficult. You will often say that, sometimes, to reduce downstream problems, upstream measures are the right ones, and at the moment the causal changes are so complex that it's really hard to communicate this. And then, of course, politicians need immediate results – that is why they always need to have some really short-term measures in their strategies because they need to show some short-term effects. But, then again, if you do something here [upstream policies] – really powerful stuff – it will sometimes be a very long time until we see the results.* (Tone Poulsson Torgersen)

> *For most of the determinants we are talking about, it's going to be a generation before you get these health outcomes. And it's not realistic to be thinking of an intervention and being able to evaluate what is going to happen a generation later. So if we are going to do evaluation successfully, the narrative that we have to get right is the causal pathways, the step-by-step factors that will lead to that health outcome eventually. And that ties in with the approach we have taken in terms of identifying intermediate outcomes and then evaluating those, but being clear that they are part of a narrative that leads you from the early years, step by step, towards those health outcomes. But we have to unpack that sequence and have a narrative that supports that unpacking.* (Peter Goldblatt)

3.3 Linking policy events to outcomes and indicators

A further challenge is that it is difficult to make links between policy events and outcomes relating to health (or the social determinants of health). There is therefore a risk that the impact of universal policies on health-related outcomes is underestimated. It is also important to recognize policies for which there is no evidence available regarding their effectiveness, and not to mistake this for evidence that these policies have no impacts on health.

> *For many public policies – and particularly in relation to upstream social determinants of health and health inequities – health is often not the only reason and primary outcome. That is often underestimated, or it's certainly not mentioned very much as a major challenge to evaluating the effects of policies on health. We risk undervaluing through our evaluations the purpose and nature of public policies. What we quite often do is to assess the evidence in relation to the potential for upstream policies to improve health, and we don't find very good evidence, but it is an issue of 'evidence of absence' rather than 'absence of evidence'.* (Mark Petticrew)

What are the most appropriate outcome indicators to select when evaluating the effects of universal policies? Sometimes health indicators may be appropriate, and they are clearly most important when trying to demonstrate that a specific policy has an effect on health. In many instances, however, upstream indicators will be appropriate for policies which aim to improve the social determinants of health. In the absence of data on health or social determinants-related effects, intermediate indicators have a vital role to play in assessing how appropriate universal policies are for achieving their aims, although there were contrasting views on this at the meeting.

> *Do we always have to try to measure the health effects? … [I]n the Norwegian strategy we don't have any downstream or health targets on diseases or child mortality and things like that, because if you set targets downstream you will very often also increase the push of the policies downstream. So, we tried instead to set objectives under the different determinants, and make our health politicians see them as health indicators as well. We could follow that line of argument – that health policy interventions that alter the distribution of health determinants are very valuable, without making the link every time to health.* (Tone Poulsson Torgersen)

> *I actually want health outcomes. If I am going to argue for a policy on the basis that it promotes health, or reduces health inequities, I want to know that it is doing that, so I still want health outcomes. But I also need to look at whether we are seeing the intermediate changes that one expects to happen, and are indicative of the changes in the system that we are trying to achieve … We need to measure intermediate outcomes and the outcomes that are conceptualized at those multiple levels – not just aggregates of individuals, but we actually conceptualize outcomes at proper levels. An example is that the notion of empowerment is different [depending on] whether you are talking about individual, organizational or community empowerment. And you can have an empowered community, without empowered individuals, and you can have empowered individuals without an empowered community. So we need to conceptualize our outcomes at the appropriate level.* (Alan Shiell)

3.4 Dealing with complexity

Many universal policies are highly complex initiatives or services. Understanding the impacts of complex interventions involves tracing and unpicking multifaceted causal chains and intertwined causal webs – this is necessarily a very complicated task. It can be difficult even to establish where interventions start and what they consist of, even before attempting to assess their impacts.

> *Any of these policies are very complex interventions, they involve very complex causal webs; this poses challenges … And that may be fine – actually you don't have to treat everything as being highly complex – it depends on the sort of answer that you want. But in fact, the sorts of interventions that we end up evaluating in real life are like this: 'Something is happening, I am not quite sure, can you see it?' And eventually it fills in and actually something happened. 'But wasn't it already going on before the intervention? And I am not quite clear because there are lots of gaps but it did seem to go down afterwards. … This sort of reality I think is much less easy to deal with.* (Mark Petticrew)

3.5 The "transferability" problem and the importance of context

It is vital to take account of the contexts in which interventions are implemented, when seeking to explain outcomes. The outcomes of initiatives are contingent not only on the characteristics of the population receiving that intervention, but also on the specific context in which they are being implemented. The organizational context is also an important aspect of the intervention.

> *Obviously we are interested in the context in which these interventions are deployed, and so that's not just the target population characteristics, but the characteristics surrounding them, in terms of environments – whether those are physical environments or policy environments. We would like to also monitor the effectiveness of the partnerships that are being developed and the level of intersectoral and interdisciplinary activity.* (Beth Jackson)

Each region or country has its own unique social and historical context, which means that different approaches may be needed to reduce inequities and to improve health in different areas. The consequence of these historical differences is that the findings of evaluations – and specific policies themselves – may not necessarily be transferable to other social or political contexts.

> *There are a number of historical differences in the social and political histories of different countries and different parts of countries, which suggests you need to adopt different approaches across countries, both to improve their overall level of health and to reduce the inequities within those countries, and it's that evidence that needs to be put together.* (Peter Goldblatt)

> *In relation to NDC evaluation,[2] there will be differences in the interventions, but there are also differences in the context within which the interventions were implemented. And one element of that is the people; another element is the places, and the history of places. These NDC areas are really diverse: some of them have a history of heavy industry in the north-east and north-west of England; some of them have had no obvious single industry in the past. So they are quite diverse in terms of their industrial labour market history. The other thing we are trying to do [in the evaluation], which is a bit trickier, is to try to get a handle on where they were going before the NDC. So were their unemployment rates going up or down or had they been stable for the decade before the NDC [started]? And this is proving to be more difficult.* (Jennie Popay)

These challenges in understanding which aspects of findings from the evaluation of universal policies can be transferred to other policy situations can be described as "the transferability problem". Two key questions underlie this:

1. To what extent can any initiative be exactly replicated in another setting?
2. To what extent can the impacts of any intervention be reproduced in another context?

Perspectives on the nature of the transferability problem depend on understandings of internal validity, which in turn is rooted in philosophical assumptions about the nature of reality, and how science claims to know or to assess that reality – all of which are contested. It may well be adequate – or "good enough" – to be aware of the limits of each evaluation and to exercise caution when transferring findings to other settings, particularly to other countries.

> *Where there is clear reason to assume that what has been observed in one place (where the primary studies were originally done) will apply equally well in another, you avoid the transferability problem. What, again, is very striking is that all of the evidence-based medicine approaches seemed to focus on the question of internal validity: to what degree can we be certain that the relationship between the dependent and the independent variable(s) is the real one? And if only we can eliminate all that bias along the way, we will get to that underlying reality. But, in fact, the notion that there is an underlying reality between those two things is questionable. Of course, that still holds up if the intervention itself has been so well described that you can replicate it, and people who know nothing about the original work can do it without knowing much about it. But that turns out to be a rather rare situation.* (Mike Kelly)

[2] NDC (New Deal for Communities) is a United Kingdom policy initiative aimed at reducing place-based inequalities through coordinated action at the local level.

3.6 The mismatch between research and policy time frames

The relationship between researchers and politicians has been described as a "dialogue of the deaf". Different dimensions of two key questions can be considered:

1. What impact does research evidence have on policy-making?
2. What impact does politics have on the production of research evidence?

The mismatch in time frames between the relatively slow pace of (even rapid) evaluations and the much faster pace of political decision-making can be a barrier to the use of research evidence by policy-makers. Politicians frequently need to respond rapidly to social problems, and there may only be limited existing evidence available from the evaluation of past initiatives that is applicable to the current policy crisis point. From a researcher perspective, the ideal way to respond to a policy challenge would be possibly to conduct a systematic review, then to design an intervention, to pilot and evaluate this initiative, and then to roll that intervention out over a larger area – perhaps nationally. This process would take years, and politicians are unable to wait when faced with a political crisis which must be tackled – often immediately. This means that the impact that existing research has on policy development may be limited if politicians do not take the time to interpret and to synthesize the relevant evidence that has already been produced. It also means that researchers typically only assess the impact of interventions after they have already been rolled out across a wide area, and those initiatives may even have reached the end of their policy lives as evaluation findings are produced. Although research should have a role to play in decision-making, efforts are required to change this "dialogue of the deaf".

> *A group of senior doctors and academics were all urging Kenneth Clarke (then Minister of Health for England at the time of the 1990s NHS [National Health Service] reforms) to evaluate these reforms, saying 'This is a huge natural experiment', 'You are being incredibly brave doing this without an evaluation'. Now, Mr Clarke didn't buy it and said 'Well I am sorry, ladies and gentlemen, but you don't understand how these things work in government. If you think we are going to set up a three-year evaluation of this in Wigan, and that we are then going to wait until we get a report which says "Well on the one hand, and on the other hand, and all things being considered and all things being equal, this might be the way forward", I am telling you this isn't the way the world works.' He simply said "This Government has decided to do these things, we have made the best possible judgements that we can, we have analysed it as carefully as we possibly can, and this is what we intend to do".* (Richard Alderslade).

> *How are we going to say [to policy-makers] 'Well, actually, scientific evaluation, analytical, thoughtful interrogation of these issues really does have something*

> *positive to say to you, but we are not going to ask you to wait three years for the results'?* (Richard Alderslade).

Even if evaluative research evidence is produced within political time scales, it is also potentially problematic for decision-makers that the findings of many evaluations are equivocal. It is difficult for politicians to know how to interpret and apply research evidence which provides only evidence of weak effects, or which presents seemingly contradictory messages about impacts and non-impacts in relation to different dimensions of the intervention being studied. It is likely that these issues of interpretation and applicability contribute to politicians' lack of interest in evaluation evidence, and engaging politicians in evidence for the impacts of universal policies is a key challenge.

> *Political engagement is going to be a significant and difficult part of this. My observed experience of having worked very closely with government is that the closer you get to it, the more chaotic it is. And partly that chaos comes from all the competing interests that operate, because as policy is allegedly being formed and it develops, all sorts of competing interests try to influence[it] to varying degrees. But that in itself is not a logical process because I am often struck by the degree to which the policy-makers and the politicians haven't got the slightest idea of what to do beyond a very general level of a desire to do something and to be seen to be doing something. The social and economic problems are of such degrees of complexity, perhaps it's inevitable that that is what it is like. But the reality when you get very close to it is far from rational and it's certainly far from knowledge driven. We need to find a way of capturing that. There is a very real danger that if we over-intellectualize it, and describe it in over-rational terms, we will chase after something that isn't really there.* (Mike Kelly)

> *I have been involved with the Department of Health to try and make the case for a new approach to the design and implementation of complex community-based initiatives. But, of course, this makes the very large assumption that ministers and their advisors do have a genuine interest in creating knowledge, to make a real contribution to the evidence base, and that may not be the case.* (Ken Judge)

A further issue for politicians' interpretation of research is that of "fig-leaf evidence", where the existing evidence base is interpreted in an expedient way, to justify political decisions which must be made.

> *Sometimes politicians use fig-leaf evidence – where you have a policy you want to introduce and you are looking for the evidence to support it, and then the wider evidence is neglected or ignored or misinterpreted. There is an interaction there between the evidence and policy.* (Margaret Whitehead)

In many countries, it can also be difficult to secure research funding to evaluate universal policies, perhaps because of the lack of political interest in the findings of such evaluations. This is particularly an issue at a time when universal policies are being eroded across Europe.

> *The practical issue, coming from a country [Sweden] in which the national Government is no longer really interested in these issues, I find it really difficult to find funding for these kind of activities. And for the last round in the Commission on the Social Determinants of Health we had a different government – I am not sure they were more interested really – but they actually put up some funding for the NEWS project, which was helpful.* (Olle Lundberg)

A further challenge for the evaluation of universal policies and natural experiments is that changes at the political level may interfere with evaluation study designs, for example by bringing initiatives prematurely to an end. This premature conclusion of interventions disrupts the generation of longitudinal data on changes, and there is then a risk that researchers will try and assess impacts at a very early stage, to preclude premature closure of the initiative. This pragmatic solution, however, may be counter-productive if it leads to premature findings.

> *If the implementation is in the control of politicians then they will come along either at the beginning or at the end and mess up your randomized control design. They will mess up the purposes of the intervention half-way through. Or, if you intended it as a pilot, they will say 'No, we visited a pilot site, it is so wonderful we must roll it out to the whole country tomorrow', before you have finished your evaluation and had a chance to compare the pilot sites with the areas which don't have the intervention. And so your entire evaluation is ruined.* (Peter Goldblatt)

> *We need to think how we can design the evaluation process in a way that will be robust to politician reality, as well as to the needs relating to strengthening scientific evidence about social determinants and effective action.* (Peter Goldblatt)

3.7 The difficulty of engaging decision-makers from other sectors

Many of the actions required to improve the social determinants of health are implemented outside the health sector, so it is therefore important to engage decision-makers from across the public and voluntary sectors in evaluating universal policies. Despite this, it can be difficult to engage policy-makers in intersectoral evaluations:

> *In Poland we have a sort of an Orwell 1984 problem, in which certain things are named but other things are behind them. Mention of 'Equity' is a key example. There are some principal documents, such as the National Health Programme for 2006–2015, in which equity is mentioned only once there in general [terms], in an introduction to the report. If you look at the objectives and you look at the particular aims of the programme, however, there is nothing more on equity. So, although there is an approved national strategy which contains certain topics,*

you don't actually find anything about equity in the priorities. (Maciek Godycki-Cwirko)

In Slovenia, social solidarity is still strong, so in this respect we would align much more with Scandinavian countries, but for a Minister of Health it would be still very difficult to argue the equity issue only from a social solidarity perspective. He would also have to bring some economic arguments slowly into it and I think in certain eastern European countries there is even more pressure to employ the economic argument... in many of our environments, it is not yet the culture to see health as a part of development, or social determinants as part of a health agenda (Tatjana Buzeti).

3.8 The need to prioritize interventions for evaluation

While evaluations are clearly of great importance in understanding the impact of population-level policies, it is vital that researchers decide carefully which interventions are worthy of evaluation. In the context of finite resources to conduct research, there is a risk that evaluations are carried out which do not represent a sufficiently valuable opportunity for policy learning. Similarly, there is a danger that researchers will always want more evidence, and therefore continue to study initiatives when perhaps sufficient evidence is already available. Researchers therefore need to focus closely on the "evaluability" of initiatives and to consider conducting fewer – but more useful – intervention studies.

We very rarely think about the 'evaluability' of interventions. There is a real risk that what we end up doing is reacting to policy needs all the time, and saying 'We can evaluate it'. As researchers, I think actually we don't think enough about prioritizing some of the evaluations that we do into those that are worth doing and those that we could do but they are not really worth doing. When you read all the rhetoric that is produced by academics, people like myself and others, there is an implication that everything is an opportunity – actually, that is wrong. They may all be opportunities, but we don't have to seize them all. We should very early on stop and assess the evaluability of any of the interventions that we are presented with … there are relatively scant resources in terms of people and money for actually evaluating any of these [initiatives] … We should be doing fewer, but better evaluations. (Mark Petticrew)

There is the 'Oliver Twist law of evidence', in that researchers will always ask for more evidence no matter how much there is, without considering the costs and benefits of doing so. And we tend not to consider issues of prioritization: how the evidence might be used, whether it's needed and how will we know when we have enough. (Mark Petticrew)

Will we ever have enough evidence? I think the answer is 'no'. This is because evidence on its own – no matter how good – will never in itself be sufficient for

> *the kind of real-life decisions which we are interested in talking about*. (Mike Kelly)

3.9 The tension between "robust" and "good enough" evidence

There is a tension between the need to generate robust evidence that is considered convincing by decision-makers and the difficulties associated with assessing the population and differential impacts of universal policies. In this methodologically challenging context, there is arguably a need for acceptance of evidence that is "good enough", which is a pragmatic response to the "real-world" policy situations faced by researchers and politicians.

> *The issue will often be about – in second-best situations – putting together convincing and plausible narratives, and actually skating over the realistic evaluation bit.* (Ken Judge)

Part of this argument for the acceptance of "good enough" evidence is the recognition that the claims that can be made for research evidence are limited, and that no scientific evidence can claim to capture and account for every dimension of complex social interventions. The belief that epidemiological studies capture observable truths can therefore be described as a 'fundamental fallacy'.

> *There is a fundamental fallacy at the heart of epidemiology; [the notion] that what is being described is an observable truth, rather than a representation of reality that can best be described by the available scientific tools to hand. And of course that is true of all sciences; it is your best attempt to describe the phenomena in front of you as they appear to be, as objectively and as honestly and with as much integrity as you possibly can. But always with the proviso that it can only ever be tentative – it is not a representation of an underlying, unsullied relationship between two variables … what we do need to do, however, is become rather more robust in our arguments about the nature of what data are, what data represent and what the limits of the data are that you can work with.* (Mike Kelly)

Section 4: Current promising approaches to evaluation and gaps/refinements needed

It is clear that further evidence is needed in relation to the impacts of universal services and policies. There is currently only a limited body of robust evidence of the outcomes associated with universal policies because of the multiple technical challenges related to evaluating those outcomes. This section sets out some diverse and innovative approaches to assessing impacts on health and the social determinants of health, both at the population level, and also in relation to health equity effects. It also outlines those methodological areas in which further development is required.

4.1 How might universal policies be evaluated?

There are a number of promising approaches which have been or could be refined for the evaluation of universal policies. It is important to learn from both policies and services which have had a positive impact on improving health, and also from those which have had limited effects or have even caused harm.

Exploiting natural policy experiments

Because a strict experimental design is generally not possible in the evaluation of universal policies (for the reasons given in section 3), the search is on for natural policy experiments that can be exploited for evaluation purposes. By "natural policy experiments" we mean naturally occurring policies or interventions introduced in one or more countries or in different places or different time periods in the same country, over which the researchers have no influence in terms of who is (or is not) exposed to the intervention. The researcher has no control over the implementation of the intervention; they can't control who receives it, or how much of it they receive. Unplanned, uncontrolled natural experiments happen all the time, often without any real prospect of anyone learning anything substantial regarding the effects of those interventions on health or health inequities. For example:

- building new roads
- privatization of public services
- new employment policies
- crime prevention/reduction measures
- new housing policies
- health service reorganization.

Because of these problems, natural experiments that are evaluated are often ones where the effects of the intervention are significantly wide reaching and involve short causal chains. It is important to note that a "natural experiment" is not a type of study design, but rather is a property of the intervention itself, and no specific evaluation design is implied. The appropriate choice of study design for the evaluation of natural policy experiments depends on the research question. If we want to know about the outcomes, then some sort of observational outcome evaluation is appropriate, such as interrupted time-series or a simple before-and-after study. Equally, if we want answers relating to processes and mechanisms or acceptability, we might carry out a completely different sort of evaluation, which could involve quantitative and qualitative methods.

> *'Natural experiment' is a misnomer anyway, and in the literature it is often used as if it was a study design. A natural experiment isn't a type of study design – it's not in any way an experiment, it's not a method of evaluation. It's a metaphor. The words 'natural experiment' actually refer to the intervention … I think that is quite important. Because we talk a lot about natural experiments as if they were only these big uninterrupted time-series studies, and given that 'natural experiment' refers to the intervention and is not a study design, there are no specific implications for evaluation. You can evaluate the intervention, depending on what sort of answer you want, so the evaluation that one carries out depends on what sort of answer one wants. If we want to know about the outcomes, then some sort of observational outcome evaluation is appropriate: that could be an interrupted*
> *time-series study, a very simple before-and-after study, or a controlled before-and-after study. Equally, if you want answers relating to processes and mechanisms or acceptability, you might do a completely different sort of evaluation, which could involve quantitative and qualitative methods.* (Mark Petticrew)

The use of the term "evaluation" can also be questioned, in this context.

> *The question was raised as to whether the word 'evaluation' was the right word to use when we were looking at natural experiments. On the other hand, very valuable insights can be drawn from natural experiments. And some of those insights are quite depressing, but some are also optimistic in terms of how we can order things better. I think there is huge value in learning from experiments that have failed as well as those that have succeeded. And that I think is partly highlighted by the question 'Does it matter if the richest in any society have the best health, or is it only a problem that the poorest have poor health?'* (Peter Goldblatt).

Natural experiments can therefore represent an important opportunity to extend the existing evidence base for universal policies (see Box 4.1 and Box 4.2).

Box 4.1: Using econometric methods to assess whether the impact of economic crises on health differs in different European Union countries

Methodological lessons learnt from a study comparing the effect of unemployment changes on changes in mortality in 26 European Union (EU) countries between 1970 and 2007.

Brief overview of study

To understand how the current economic crisis may affect population health, and if any of these effects can be mitigated through government policy, a retrospective examination of the impact of economic changes on mortality rates between 1970 and 2007 was conducted. Unemployment was used as an indicator of economic difficulty. The main findings were that every 1% rise in unemployment rate was associated with (a) a 0.79% rise in suicides for people aged under 65 years; (b) a 0.79% rise in homicides; and, contrastingly, (c) a 1.39% reduction in road-traffic deaths. Sweden, however – which has one of the best-resourced labour market protection systems in Europe – de-coupled this association; despite the banking crisis in the 1990s, when unemployment rates rose by about 10% within a very short period of time, there was no correlation with suicides rates during the same period (see Fig. 4.1.2).

Fig. 4.1.1: Associations of a 1% rise in unemployment with age-standardized mortality rates, by cause of death, in European Union countries, 1970–2007

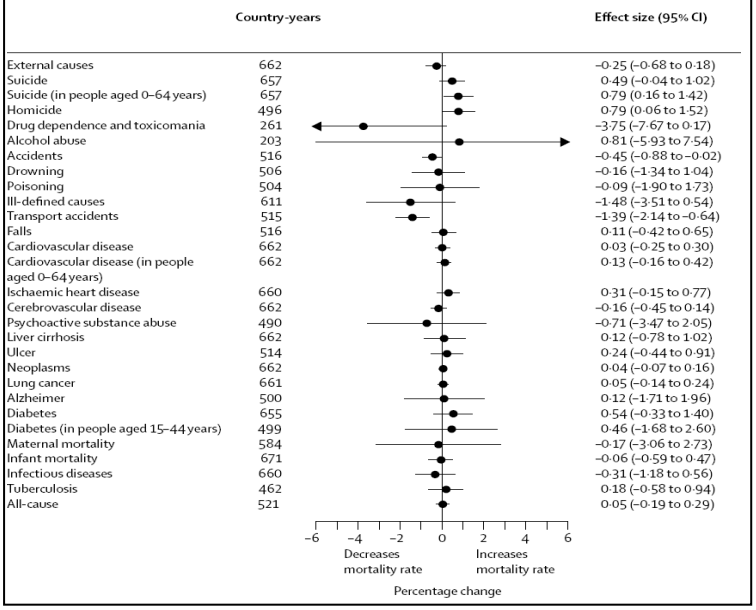
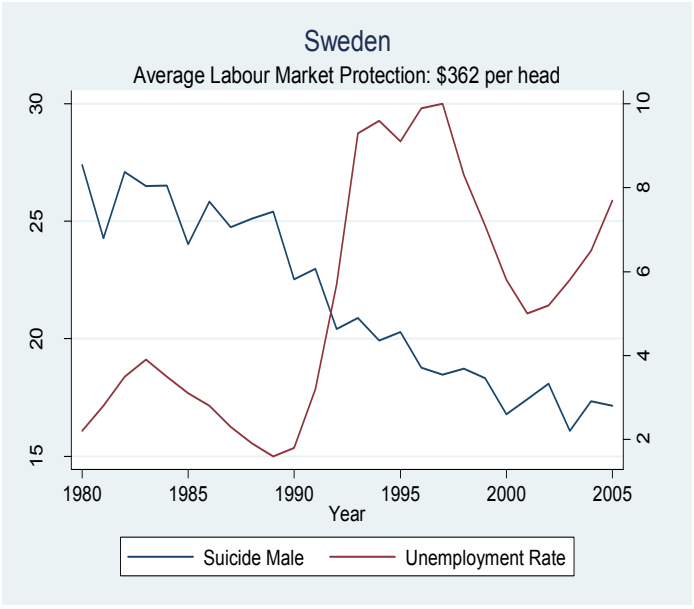


Fig. 4.1.2: Suicide rate and employment rate against years, Sweden, 1980–2005



Methods used

Data sources: Data were collected from the World Health Organization (WHO) European Health for All database, International Labour Organization (ILO) Key Indicators of the Labour Market, gross domestic product (GDP) data from the World Bank World Development Indicators (from 2008) and the Organisation for Economic Co-operation and Development (OECD) Health Data (also for 2008).

Statistical analysis: To capture the pace of change in unemployment, a comparison of changes in unemployment relative to changes in mortality was made. Deviations in unemployment rates from their average rate of change were assessed and specific periods of mass job loss (0.3% rise in unemployment in a fiscal year) were identified. A fixed-effects approach was used to isolate the association between unemployment and mortality. This method also allowed the examination of how the health of a population changes with respect to its existing health, as well as the country's ability to respond.

Commentary: Areas for methodological development

- Although increases in inequity between countries (dependent on their level of social protection) was evidenced, the health equity effects of the recession within countries is not known, as only aggregate, country-level data were available. Availability of and access to within-country data is required in order to measure the differential impact of universal policies on health inequities. ***"It is not a given that it is always the poor that are hit worse; it certainly also depends on policy and how policy responds, so the post-crisis distribution of income may not necessarily be worse than the pre-crisis one."*** (Marc Suhrcke)

- An agreed quantitative measure of the generosity of the welfare state is required. This study used spending data as a proxy for this measure; the generosity of the welfare state may appear very different to the picture we obtained from spending data. ***"In the Thatcher years, unemployment benefits rose sharply. This is hardly a sign of increasing generosity of the welfare state in that period, so this is an area we need to address."***
(Marc Suhrcke)

- The use of methods from other disciplines – notably economics and the application of econometric approaches –is essential is developing tools which can measure the differential impact of universal policies on the social determinants of health and health inequities. In addition, in order to conduct systematic reviews in the future, a consistent approach to the use of econometric methods would be essential.

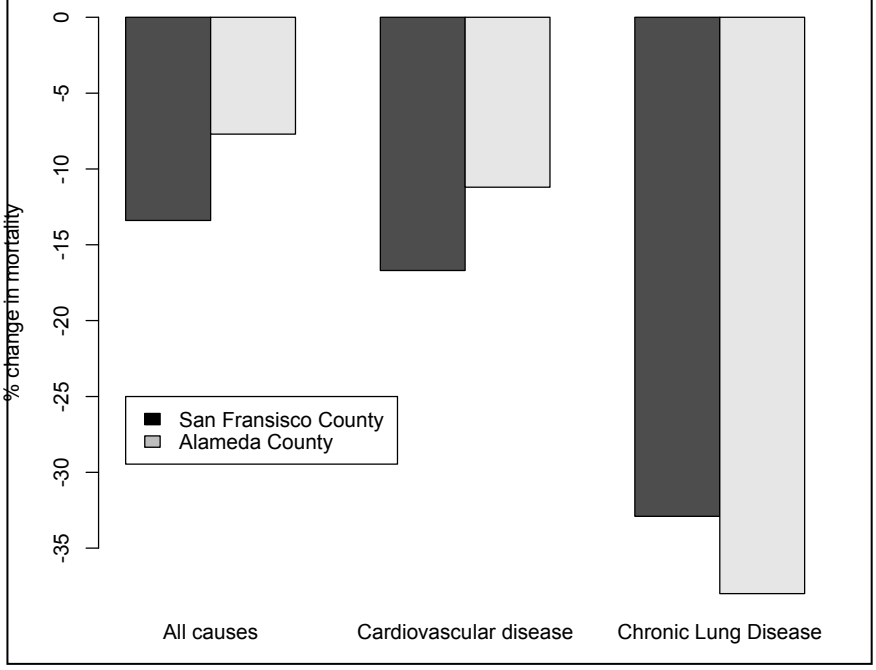
Sources: Stuckler et al., 2009; and presentation by David Stuckler and Marc Suhrcke at the meeting.

Box 4.2: The effect of the 1974 fuel crisis on mortality: A natural experiment

In 1973–1974 the Organization of the Petroleum Exporting Countries (OPEC) countries pushed up the price of oil to try to influence American policy on supporting Israel in the Yom Kippur war. Oil prices rose from US$ 3 to US$ 12 per barrel, and provided an opportunity to test the hypothesis that a decrease in car exhaust fumes would have a beneficial effect on health.

In the first three months of 1974, gasoline sales reduced by almost 10% in both San Fransisco and Alameda Counties, in California. In a simple before-and-after study, the rates of all-cause, cardiovascular disease, and chronic lung disease mortality reduced dramatically (see Fig. 4.2.1).

Fig. 4.2.1: Percentage reduction in mortality rates in two Californian counties 1973–1974



This finding was not explained by the usual seasonal secular changes in mortality, and further analysis of weather and air stability data, air-pollutant levels and patterns of influenza mortality further supported the link between air pollution and adverse health outcomes.

The results of this natural experiment suggest that policies that lead to a decrease in exhaust fumes will have a beneficial effect on health.

Sources: Brown et al., 1975, and presentation by Mark Petticrew at the meeting.

It is also challenging to assess the impact of natural experiments because they are frequently not discrete temporal events. Instead, the nature of their implementation and their effects may fluctuate over time. This means that it is very difficult to assess prospectively changes in outcome measures from a baseline. It is therefore important to design evaluations which avoid these simplistic assumptions.

> *Quite often on the evaluation side we slip into intervention thinking, and we assume that what we are looking at through these natural experiments is quite discrete and rapidly introduced on/off interventions, which can be easily described, as opposed to more gradual changes over long periods.* (Mark Petticrew)

Multiple evaluation methods – both quantitative and qualitative – can provide a range of holistic insights into the different aspects of the success of an intervention. Evaluations can seek to measure the impacts of universal policies, and studies can also generate valuable process data on implementation strategies and participants' experiences of universal services. These process and experiential data can be crucial to understanding the impacts of universal policies. Sometimes the findings from different evaluations can appear to contradict each other, and in those instances it can be challenging for decision-makers to reconcile these seemingly contradictory findings.

> *[Canadian funders] are encouraging projects to incorporate in their evaluation plans different kinds of evaluation processes – participatory, mixed method, equity-oriented and indigenous approaches to evaluation. We would like them to collect data to support the use and adaptation of the intervention, so that we can identify some of the core factors (that appear to be at work in the intervention) that might remain somewhat stable as those interventions get scaled up and disseminated to other contacts.* (Beth Jackson)

> *There is a lot of potential for integrating ethnographic methods into our evaluation processes … we need to understand who are the actors, what are the networks, the processes, the mechanisms by which at particular points in history these policies have an opportunity to emerge and take hold … Sometimes we come up with an evaluation that demonstrates the negative outcome for an intervention and it has very little to do with the intervention itself – it has quite a bit more to do with the context in which it is implemented. Until we better understand that and document those processes, we are at risk from misattributing successes and failures.* (Beth Jackson)

There may be useful econometric techniques which could be used to assess the impacts of universal policies – particularly, perhaps, labour market policies, for which their use is well established. Even using these econometric techniques, however, it may still be challenging to identify differential policy effects:

> *Econometric approaches, such as instrumental variables, propensity score matching and the like have been used extensively in evaluations of labour market policies – for instance, the impact of a minimum wage on employment in the first place, and other relevant socioeconomic determinants. Some have also tried to apply these methods to assess the impact of non-health policies on obesity; much of that work [has taken place] in the United States. But there are only a few studies – in one (non-systematic) review we carried out, we only found six studies that had made an effort to go beyond simple correlations … And in some other work we were doing for WHO, we tried to provide a modest guide on how some of these methods could be used in the areas of social determinants of health. Our focus is, however, on establishing the mean effect before we then move on, later, to trying to establish the equity effect. And even short of identifying the health equity effect on the entire population, there is a further intermediate step – to focus just on the health of the poor as the outcome variable. Can we improve the health of the poor by a certain measure, rather than looking at how health shifts across the entire socioeconomic distribution?* (Marc Suhrcke)

Although qualitative studies have much to offer – particularly in terms of assessment of the experience and process of implementing interventions – it can be challenging to integrate qualitative research into policy analyses alongside quantitative data. Policy ethnographies combine qualitative data collected using interviews and focus groups with observation, analysis of policy documents, and perhaps also secondary analysis of qualitative and qualitative data available. These policy ethnographies represent a useful method for yielding holistic, multifaceted insights into policy successes and failures.

> *We need better ways of integrating qualitative studies into policy assessment. And the use of policy ethnography – and meta-ethnography in the case of synthesis – I think would be very important.* (Margaret Whitehead)

> *Policy ethnography is needed to try to understand the policy process behind how those decisions are made, and trying to understand policy-makers – namely, their own theories of change. Policy-makers might also engage in 'symbolic politics' – the way in which the 'words can succeed, but actually the policies fail' (according to Edelman, 1971 and 1977).* (Mark Exworthy)

Evaluative research that seeks to ascertain the quantitative impacts of universal policies needs to ensure that the study is sufficiently powerful to capture the effects of those policies.

> *What might be important is assessing the size of the population, and the differential exposure, as well as the size of the effect we are trying to measure. And it may be possible to set up guidelines, based upon guidance that exists around subgroup analysis for trials. So it's important to specify in advance the size and direction of the effect we are expecting to find, so that we are clear that we will find an effect. Is it reasonable to think, for instance, that changes in the [National Health Service] NHS will quickly lead to differences in mortality? And is it reasonable to think we will pick that up?* (Rene Loewenson)

While it may seem daunting to evaluate population-level policies, problems could potentially be reduced if evaluation methods were planned alongside the interventions, from the inception of those policies. Interventions need to be better designed in order both to facilitate their evaluation, and also to influence policy and practice. Evaluation needs to be "designed in" and this is a key challenge for the future.

> *We need properly designed interventions – designing them so that they can be evaluated, and so that evaluation is built in from the start; appointing the evaluation teams before you appoint the people to do the actual implementation.* (Peter Goldblatt)

Retrospective analysis: the Inspector Morse/resilience approach

A retrospective approach (which was referred to as the "Inspector Morse" method or the "resilience" approach) can yield useful insights into the outcomes associated with universal policies by means of a retrospective analysis. Although prospective studies are generally considered to be more robust, retrospective analysis can form the basis for efficient and effective policy analysis by tracing impacts backwards from the outcome at a given point in time, in the direction of the policy being studied.

> *Where the evidence-based medicine paradigm works well is with simple downstream interventions. The evidence will tell you about the relationship between* x *and* y*, with confidence intervals if you are lucky. But the evidence will say nothing, more or less, about the intermediate steps along the way … There is a really important causal problem lurking behind this. I eventually reached the view that we were not working with a nice Newtonian causal model, that if you can get here – to* x *– somehow you should be able to predict (in the Newtonian language of physics) the outcome here – at* y*. The assumption very often is that this is what we are doing. But in public health we do something rather different – we seldom move from* x *to* y *in that way. We go backwards: we have our question here, our outcome* y*, and we ask ourselves a series of questions about what preceded it, in a forensic way – the Inspector Morse approach. When Inspector Morse finds the body, he doesn't say 'Well, what randomized control trial (RCT) led me to predict that that body might be there?', but rather, he begins a backward search into the evidence relating to the preceding conditions.* (Mike Kelly)

> *This is similar to the 'Resilience' approach to evaluation, in which the starting point is the outcome. We look and see which populations or areas are doing better than expected for their degree of deprivation – in other words, they are exhibiting resilience in the face of adversity. So we identify the outliers, and then we try to trace back, retrospectively – with an 'Inspector Morse' hat on – to what has influenced their trajectory. Why have some had a more positive trajectory,*

> *taking their socioeconomic circumstances into account? Many natural experiments need retrospective evaluation of some kind because they have been under way for some time – in the real world, there isn't the opportunity to trace impacts prospectively.* (Margaret Whitehead)

Comparative historical techniques of data analysis have considerable potential when considered as part of this retrospective or reverse causality approach (see Box 4.3). A key advantage of a comparative historical approach is the ability to examine cultural differences, and the impact that these have on policy implementation and on specific outcomes, many of which may be culturally contingent.

> *And in fact, maybe you could do no evaluation at all; you could do a historical analysis of natural experiments – a comparative history approach, looking at geographical and cultural variations between countries. And these can be seen as natural experiments of history (Diamond & Robinson, 2010), which is an entirely appropriate approach that has nothing to do with epidemiology – it's a comparative history.* (Mark Petticrew)

> *On the subject of causation, I think that we really ought to take very, very seriously this idea of retrospective understanding – cause, as it were, in the past … Thinking about the way comparative historians might look back, and we probably just have not really begun to explore the potential for that. And yet we know the history of place and the history of cultures are such important determinants of the features that we are trying to understand. And I think there is something to be said for a more scientific look at the way we could do that – and learning from the historians, archaeologists and others who piece together things from fragments of facts that are there on the specimen tables in order to do so.* (Mike Kelly)

Many researchers who seek to assess the impact of area-based interventions consider using case-control designs, because these can produce robust evidence of policy impacts. It can be very difficult, however, to identify control areas if seeking to use a case-control design, because – in the United Kingdom – all disadvantaged areas have undergone numerous interventions in recent years. It is suggested that the reverse causality approach may offer a means of dealing with this problem by enabling researchers to examine retrospectively the area-based changes that both preceded and followed the policy being studied.

> *One of the challenges to evaluation is 'burn-out regeneration' for these areas: it's impossible to find a control if you are looking at very disadvantaged areas, because they have all had something going on in them – certainly in England, and I think probably in western Europe [as a whole] – so it's a real struggle to find a control. And what the NDCs have is comparison areas, but they don't know what went on in those comparison areas over the period [in question]. So, some of the early analyses we have carried out – for example, on education – show that the NDC initiatives did improve the performance in educational terms. But the*

comparison areas improved to the same extent – and so we don't know whether that was because some of the same stuff was actually also going on in those comparison areas, even though they weren't NDC areas. So, I think this 'Inspector Morse'/resilience approach is a neat way of getting round that. (Jennie Popay)

Box 4.3: Using historical analysis to assess the differential equity impact of a universal intervention on child mortality in Stockholm

Methodological lessons learnt from a study examining the effect of universal improvements in sanitation and access to clean water on inequalities in child mortality in Stockholm from 1878 to 1925.

Main findings of the study

Improved water and sanitation was associated with a decline in overall and diarrhoea-related mortality in children aged under 2 years. Socioeconomic differences in diarrhoea-related mortality disappeared (Fig. 4.3.1). In contrast, the socioeconomic gradient in overall mortality rates persisted (Fig. 4.3.2).

"If we look at overall mortality, the class gradient remains ... so something happened with this universal intervention (improved water and sanitation) which was selectively more important for some groups more than others." (Bo Burström)

Fig 4.3.1. Diarrhoea mortality rates by socioeconomic group, children aged <2 years, Södermalm, Stockholm

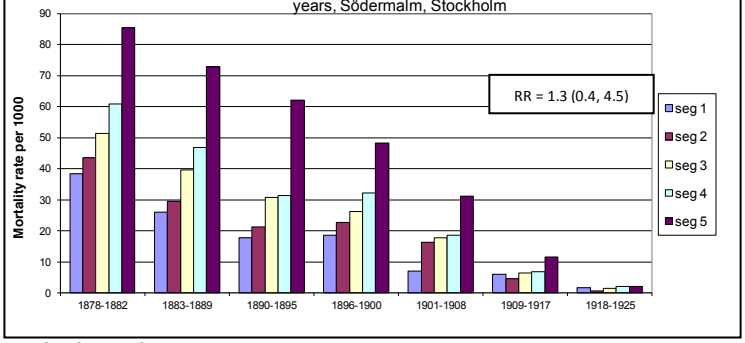
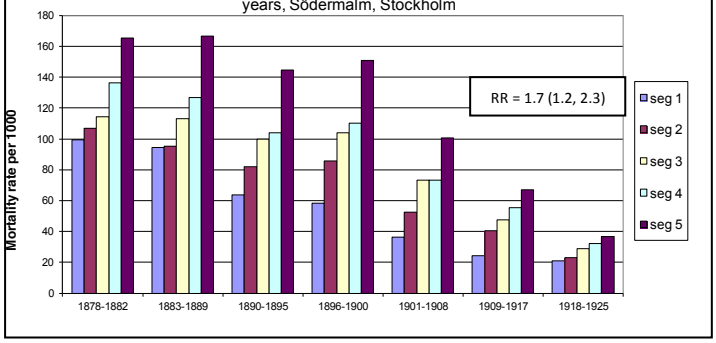


Fig 4.3.2.Overall mortality rates by socio-economic group, children aged <2 years, Södermalm, Stockholm



Methods used

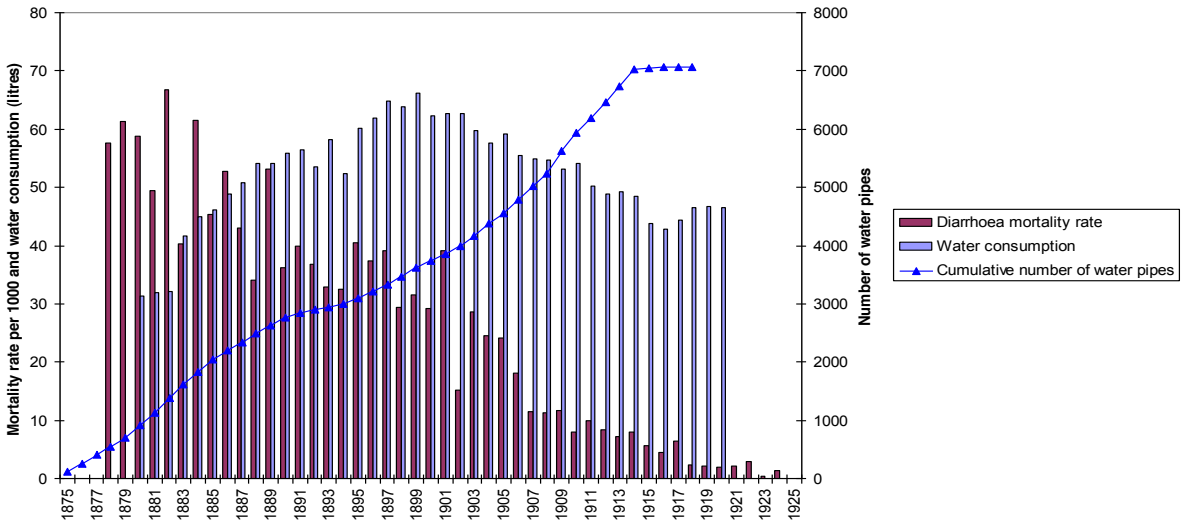
Data sources: Historical records provided information on piped water. Child mortality data were accessed from the Roteman Archives and linked to computerized death certificates originally filled out by physicians. Data were pooled into 5-year periods from 1878 to 1925. Socioeconomic group classification was based on the occupational title of the head of the household; 1 = highest, 5 = lowest.

Statistical analysis: Overall and diarrhoea-related mortality rates were calculated by socioeconomic group and time period. Cox regression analysis was used to calculate the hazard ratio of death by socioeconomic group, with socioeconomic group 1 (SEG 1) being the reference group.

Commentary: Lessons learnt and areas for methodological development

- To understand the role of improved water (Fig. 4.3.3) and sanitation in the decline of diarrhoea-related mortality rates and why there was not a concurrent equalization of mortality risk from other causes of death, historical data were accessed.

Fig 4.3.3: Mortality rate from diarrhoea in relation to daily average water consumption per person and cumulative number of new water pipe connections, 1878–1925



Other simultaneous and related interventions contributing to the decline in diarrhoea-related mortality included improved handling of excreta, better food handling, and intensified health inspection. Additionally, the majority of these interventions were universally implemented. Thus, in evaluating the impact of universal interventions on health inequalities, the **historical context** of the intervention being evaluated must be sought. It was possible to demonstrate the differential equity impact due to the ability to **link data available in Nordic countries. However, other countries may be able to make more use of their historical data and these possibilities should be explored**. The decline in diarrhoea-related mortality took **time**. Patience is required when evaluating interventions, acknowledging that the impact may not be evident within a political term in some instances.

An intervention may be more successful if implemented as part of a broader package. This is true of health care today. ***"The buzz word in Sweden today in health care is demand; if you have a need, you demand health care – and I think we go wrong in many ways [with this approach]. We need much more emphasis on health promotion, on primary prevention, secondary prevention ... And we also need to consider equality of care so as to affect the outcome."*** (Bo Burström)

From randomization to case studies

There may be a place for randomized study designs in the assessment of universal policies, in order to produce "strong" evidence which policy-makers find convincing. Some researchers continue to regard RCTs as the only truly robust form of evidence. In most instances, however, these randomized designs cannot be feasibly or appropriately used to evaluate public health interventions. This may mean that some researchers and decision-makers remain sceptical about the value of findings produced during the evaluation of public health initiatives.

> *Even with all of this going on, you can still randomize. Which is what we are going to do [in Canada]. Randomization is not clear-cut – I don't want to over-standardize, I want variation and experimentation going on, but I quite like a randomized design. Because in health promotion, I need to go up to a Department of Health and Welfare Services and say 'I have got the same quality of evidence, about social capital-building interventions, as the neonatologists have got about intensive care'. That is the fight that I face, and if I cop out of that, we are going to continue to marginalize health promotion.* (Alan Shiell)

> *I have met several people in the RCT world who I can only describe as zealots. But I think it's perfectly clear that RCTs have a very positive contribution to make to the overall enhancement of the evidence base, but that they are limited, and that there aren't many situations in which they are the necessary or the appropriate response.* (Richard Alderslade)

In some policy and academic arenas, narrowly defined research designs – such as RCTs – are considered valid, particularly for answering questions relating to the effectiveness of clinical treatments. This perception regarding what constitutes robust evidence – which originated in the health care sector – has had a considerable influence on public health. Increasingly, however, people working in public health accept that different kinds of research designs are appropriate to answer varying research questions across diverse policy contexts. Legitimate methods are considered to be those which are appropriate for the evaluation in question. Furthermore, different kinds of evidence are considered to be convincing in different policy sectors, and this is particularly significant in relation to universal policies, which seek to improve the social determinants of health. For example, case study evidence can yield valuable insights into complex social situations.

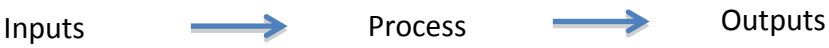
> *What we don't know enough about is the value of evidence within different sectors, and what people find useful. One of the reasons why there aren't more RCTs of housing is partly related to the different disciplinary histories, but is also because the sort of evidence that people who plan housing act upon isn't trial based. The sort of evidence that they need is actually – quite appropriately – derived from weaker study designs – survey evidence – and I think we don't know*

enough about that from within other sectors. Case studies in particular are relevant because they are very popular outside of [the] health [sector], for example in urban planning and transport. Now, some of the reason for that is that they tell those very interesting, complex narratives that include surveys and some experimental quantitative evidence, but also qualitative information is conveyed in a way that makes the whole thing useable. I think we just assume – in public health anyway – that case studies are bad, because they are not robust enough and quite often they are not transparent. But actually we don't have any really clear picture of what sort of evidence is needed, or is useful to decision-makers across different sectors. (Mark Petticrew)

I do think that we are paying quite a high price in public health for the success of evidence-based medicine, and the idea that in different sectors there are different types of evidence is a relatively recent phenomenon. Actually, the acceptance of diverse forms of evidence was quite common in the health sector in the early 1990s. And it's right that we had to move towards a more systematic, transparent way of evaluating effectiveness, but it's almost like we have gone too far. And so I think public health needs to be a bit more courageous about its challenge to the evidence-based medicine paradigm, and the limits of that. (Jennie Popay)

Insights from Complexity Theory

Standard economic evaluation approaches often assume a very simple underlying model, as shown here.



However, once we move into the notion of complexity (Shiell, Hawe & Gold, 2008; Hawe, Shiell & Riley, 2009), these assumptions are undermined. Complex systems are self-organizing, dynamic and have a trajectory dependent on previous history. They are adaptive, and can "learn" new behaviours, but also can get stuck in suboptimal positions (basins), so that change can be "sticky", and when it does occur (the tipping point) it is often dramatic (phase transitions). Feedback abounds in complex systems, and this further challenges simple assumptions. For example, with smoking bans, support goes up after the ban and so the value that people will attribute to a programme is a function of the success of that programme (endogenous preferences). Furthermore, outcomes are often multiple, and occur at multiple interacting levels (emergent properties). Thus, complex systems appear to be more like the Foresight Obesity model, which is rapidly becoming renowned as an impossibly complex picture.

Complexity is therefore not only a property of the intervention, it is also a property of the system or context, and there is a high degree of human agency

involved, which means local adaptation of the intervention is unavoidable. This means that spin-off benefits and costs (externalities) are to be expected, but their precise nature is difficult to predict. The "cohesiveness premium" (phrase coined by Popay) associated with universal policies is an example of this, since social value is a collective property, related to a sense of cohesion and connectedness.

> *A key point that emerges is that we need to stop thinking about interventions as technologies which are done to people, in the same way as I give a drug, or I do some surgery, and start thinking of them instead as events in the life of a system that can either be embedded or can be washed off, depending on how well it [the intervention] engages with the context.* (Alan Shiell)

In practice, one can use changes observed in complex systems, such as alterations in social networks or the flow of resources, as indicators that an intervention will have an effect on the desired outcome (for example, a reduction in health inequities). These changes should be conceptualized at multiple levels; for example, empowerment can be at the individual, organizational or community levels, and it is possible for an individual to be empowered without an empowered community.

The Medical Research Council guidance on the evaluation of complex interventions also contains useful insights for the assessment of universal policies (MRC, 2008).

Utilizing logic models and a systems approach

The evaluation of universal policies and services requires understanding of the causal processes which underpin interventions. It is important to understand the theory of change to build logic models and the 'Context-Mechanism-Outcome' (CMO) configurations which are associated with these. Although these have the potential to yield key insights, they can be difficult to describe and to untangle, particularly at the local level.

> *What I have frequently observed with a wide variety of initiatives is a lack of capacity (or possibly even discipline) at the local agency level, to plan effectively or to use resources clearly or intensively enough to make a real difference. Most importantly – 'Context-Mechanism-Outcome' configurations: lots of people talk about theories of change and 'Realistic Evaluation'. Developing CMOs as well as logical models in the complex environment, to make serious inroads into developing effective public health action is extremely hard work; it is a serious struggle, and people don't do it very well, and they are expected to do it really quickly.* (Ken Judge)

> *We are going to look at the NDC – if we do find an impact, and some of our preliminary work suggests there is an impact – then does it vary across types of NDC intervention? … We are trying to characterize these NDCs according to their underlying theory of change, so trying to get a handle on whether there is a way of distinguishing between them at that macro level, because that is the interesting thing from our point of view.* (Jennie Popay)

A systems approach may be useful for the evaluation of universal policies, encompassing both complexity and considerations of context, mechanisms and outcomes (see Box 4.4). The systems approach also recognizes that interventions are dynamic and adaptive, and appreciates the value that local innovation and adaptation can bring to the design and implementation of initiatives. From this perspective, interventions are conceptualized as events occurring within systems. This presents challenges for evaluators, who have to reconcile the need to evaluate an intervention (and within a context) which is fluid and constantly adapting, with the need to generate consistent data which can measure and assess changes and impacts over time.

> *I have come to a kind of synthesis between the MRC view (which was that complexity was a property of the intervention) and the complex systems view (which is that complexity is a property of the system), to recognizing that we are actually talking about both, where there is a high degree of human agency involved. Local adaptation of the intervention is unavoidable; it's not something you just permit, it's unavoidable whether you like it or not. In fact, I like it, because I think we have to adapt interventions to local context. History within public health shows that without adaptation you get ineffectiveness and spin-off benefits and costs. So we need to look for the unexpected. What the systems perspective does is to emphasize all of these, but it also emphasizes the dynamic – this notion that the system is moving over time, it has a trajectory and a movement of itself, and I think that affects the way we think about intervention.* (Alan Shiell)

Cross-national comparative policy analyses

A key way of examining contextual differences in policy impacts is to compare outcomes and to identify either similarities or ways in which these differ between countries (see Box 4.5).

> *You can look at differences between nations as continuous variables, degrees of differences in different types of policies. And looking at policy institutions, make some more room for causal influences: because you have the factor identified that you are looking at, you can specify relations, you can test relations.* (Olle Lundberg)

When conducting cross-national policy analyses, it is important to identify the principal features of different countries' approaches to social policies. This is

necessary in order to understand the higher level logic models and historical trajectories which form the policy environments in which universal policies are designed, implemented and evaluated. When examining these policy contexts, it is important to resist the simplistic clustering of countries into welfare regimes.

> *Whenever we compare countries, if you take 18 countries and collapse them to three regime types you have an even bigger small-end problem, and I am not really sure why we should do that. And it's very difficult to say which kind of cluster principles we should use, because countries cluster differently according to different principles and these differences are not stable, even in time. So should we class the countries according to their properties in the 1950s, 1960s, 1970s, 1980s, 1990s, or whenever?* (Olle Lundberg)

Further developments are required to facilitate a greater understanding of how to compare policies and governmental approaches when conducting cross-national analyses. Researchers who conduct these kinds of analyses must ascertain when it is appropriate to compare policy events, with detailed examination of interventions and their contexts required to assess whether policies are actually comparable. It may be more straightforward – rather than taking as a starting point the outcomes of specific policies – to compare the experiences of certain population groups, which are more immediately comparable.

> *We find that when we are doing cross-country analysis, many of the policies look very different on the surface, but there are underlying similarities – there are approaches that are based on a similar theory of change. Even if the way the policies are implemented, the names of the policies and the way they are described are different, you can differentiate distinct types that are based on distinct strategies and theories of change. And so we do believe that taking this approach forward needs more development, more understanding of the different types of policies, to be able to do these sorts of comparisons in a valid way.* (Margaret Whitehead)

> *The way it has worked for us in doing these analyses is by limiting the analysis to a few countries and having in-depth knowledge of these countries – because you have so many contextual factors to handle; the world is messy. But I think if you compare two contexts, it's easier to pick apart what is policy and what are other contextual factors within that. I wouldn't think of using in-depth comparison for 20 countries or anything like that, but I think when you have limited context then it is useful.* (Bo Burström).

Policy variations within regions of the same country can also be exploited effectively.

> *There has been considerable work in the United States which exploits differences in various policies across the states, comparing the outcomes of interest (for example, in education or other public policy areas) over time – using a difference-*

> *in-differences approach to the outcomes – between the states that have and have not implemented a given set of policies.* (Marc Suhrcke)

Assessing differential impacts

Even if a universal policy is applied across the whole population, it may still have a differential impact, greater for some population groups or areas than for others. Assessing policies and interventions for differential impact is imperative from an equity perspective, but has been a neglected component of intervention research for too long – partly because of the difficulties involved in doing so. This leads some to advocate a gradual, step-by-step approach.

> *Are we taking too far a step by jumping immediately at the ambition of wanting to assess health equity impact here? Should we not first set ourselves the more modest goal of trying to better assess the average impact? That is one question.* (Marc Suhrcke)

> *We try to first establish the mean effect before we then move later on to trying to establish the equity effect. And even short of identifying the health equity effect on the entire population, it is again a further intermediate step to just focus on health of the poor, as the outcome variable. Can we improve the health of the poor by a certain measure, rather than looking at the whole distribution?* (Marc Suhrcke)

Box 4.4: Using complexity and systems theory to inform evaluation

A number of large, well-designed community-based evaluations of health promotion have at best shown very weak effects, and at worst show harm – in terms of often perpetuating inequities rather than dealing with them. One hypothesis to explain this is that too many community-based interventions are theorized at too low a level (that is, the level of individual behaviour change) and ignore the high-level influences that occur elsewhere in the system. Taking an approach informed by systems theory and Complexity Theory can lead to conceptualizing and evaluating interventions in a different way. In the example below, social network analysis is used to capture the complexity of social relationships in a school in a way that has applications for health promotion research.

Using social network analysis to map the social relations of staff and teachers at a school

Social relationships can enhance or hamper quality of life, coherence, sense of personal worth, health and educational and economic opportunity, and are important outcomes for health promotion interventions. In order to quantify the degree and nature of social relationships amongst staff in a school, social network analysis was used to characterize the school in terms of the density of relationships and the connectedness of a community of staff and teachers. The approach contrasts degree centrality – the number of connections an individual has, with between-ness centrality, which is a better reflection of a person's strategic connection to the most marginal people in a community. This approach can also be used to identify strategic "champions" and gatekeepers in communities, who can be targeted in health promotion interventions.

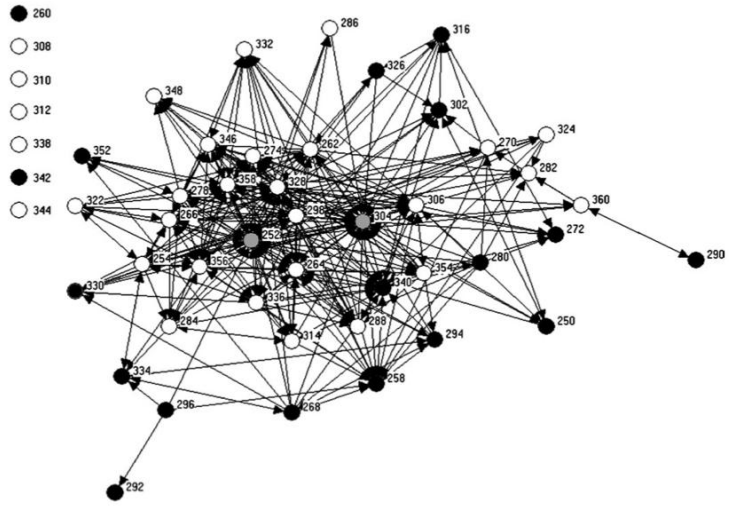


Fig 4.4.1. An Advice-seeking network. Grey = Principal and Vice Principal. Clear = teachers. Black = support staff. Who do people turn to for advice? The most central people are the Principal and Vice Principal, and the school secretary. Teachers are more central in the network than support staff. The proportion of reciprocated ties in the network is 29% meaning that most relationships are one-way

These approaches are being used to evaluate a school-based health promotion intervention aiming to increase connectedness between pupils, between staff and pupils, and between the schools in question and the community (multi-level approach). Social network analyses of connectedness are used as outcome measures of social isolation in a cluster-randomised design.

Sources: Presentation by Alan Schiell at the meeting; see Hawe et al., 2008.

Box 4.5: The Nordic Experience of Welfare States and Public Health (NEWS project)

Methodological lessons learnt from a study exploring how variations in the design and generosity of family and pension policies are linked to infant mortality and old-age excess mortality, respectively, across 18 Organisation for Economic Co-operation and Development (OECD) countries.

Main findings of the study

Increased generosity in family policies that support dual-earner families is associated with lower infant mortality. Generosity in basic security pensions versus earning-related income security pensions is associated with lower old-age excess mortality.
"It is not just spending more money; it is on what type of policy you spend the money that is important." (Olle Lundberg)

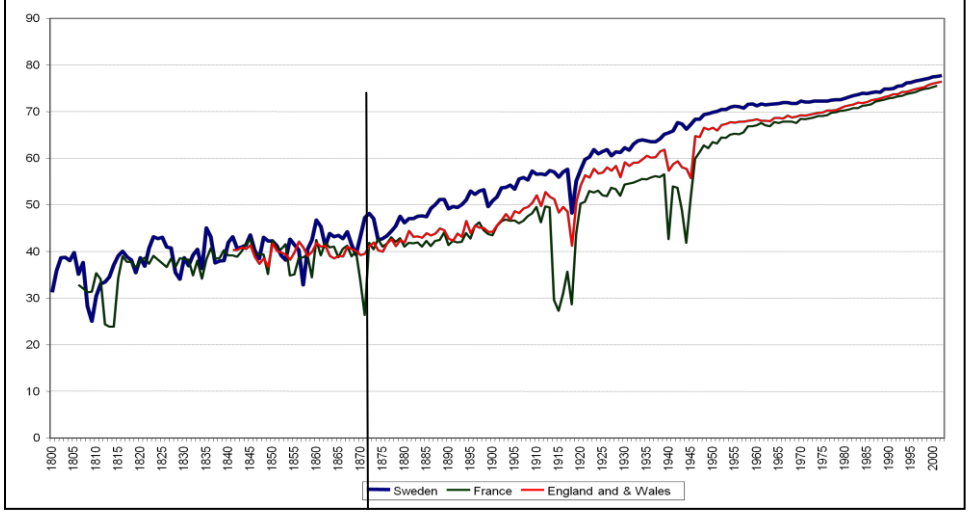
Methods used

Data sources: Indicators of policy characteristics were derived from the Social Citizenship Indicator Program. Two types of information were used in the analyses: (1) characteristics of family policies and pensions; and (2) the generosity of these policies, as measured by the replacement rate. Family policies were grouped into typologies based on their characteristics: the dual-earner model, the general family policy model and the market-oriented family policy model. Likewise, pension policies were grouped into: the encompassing model, the state corporatist model, and the basic/targeted model. The Human Mortality Database and the World Health Organization (WHO) Mortality Database were used to access mortality data. The Angus Maddison's databank, International Labour Organization (ILO) and OECD were accessed for gross domestic product (GDP), female workforce participation and unemployment data, respectively.
Statistical analysis: Pooled cross-sectional time-series analyses were conducted for pension policies relating to 18 OECD countries during the period 1970–2000 for family policies and 1950–2000. Data points at every fifth year were used, leading to 108 observations for family policy analysis and 162 observations for pension policy analysis. Fixed-effects models were used and known potential (mortality) confounders were controlled for, such as GDP.

Commentary: Areas for methodological development

- In the evaluation of universal policies, how can the contextual factors that affect the effects of policies today be captured?
  "The differences between countries go way beyond or way before the welfare state ... the results that we have today may be based on interventions we did 40 years ago, 50 years ago, or even longer." (Olle Lundberg)

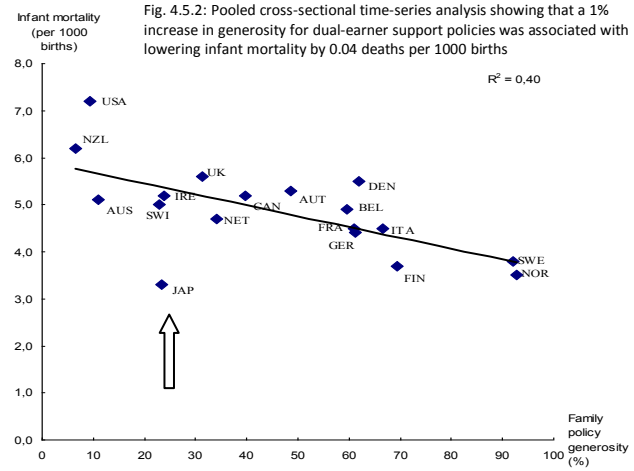
Fig. 4.5.1: Life expectancy for males, 1800–2001



"From 1800 to 2001, around 1858 the life expectancy in Sweden amongst Swedish males was higher than in the other two countries. In the mid 19th century in Sweden however, we had an alcohol epidemic of the size of Russia today, so there is no logical reason for Sweden being better off... and as you know the welfare state comes in around (1870), being built during the 19the century, so something about the country differences is not about the welfare state, as we know today, it's about something else." Olle Lundberg

In making cross-national comparisons of social policies, caution is required in categorizing governmental approaches into specific typologies. An analysis of the detail and context of the intervention is needed to ensure the appropriateness of the comparisons made and also to differentiate between the combined vs. specific effects. One way forward is to look at within-country specific policies rather than combining countries.

Fig. 4.5.2: Pooled cross-sectional time-series analysis showing that a 1% increase in generosity for dual-earner support policies was associated with lowering infant mortality by 0.04 deaths per 1000 births



"Though there is a somewhat linear trend, these clusters of countries are not behaving the same, so in my view we are better off looking at these countries individually and their specific policies rather than combining countries. For example, in the cluster with very small efforts in family policy generosity, you have huge differences, so what is Japan doing? They are doing something that is reducing infant mortality whilst at the same time having very low ambitions in family policy generosity." Olle Lundberg

Sources: Presentation by Olle Lundberg at the meeting; see also Lundberg et al., 2008b.

It can be difficult to assess changes in health inequities because of a lack of consistent measures over time – this may be a particular issue for retrospective approaches to assessing differential impacts. It can also be difficult to assess differential impacts in instances in which individual-level data are not available, and where there are displacement effects.

> *Getting a clear handle on the impact on health inequities is difficult because of the absence of long-term measures of health inequities. So that is a serious constraint for us, when we look historically at some of these issues.* (Ken Judge)

> *What does our work say about health equity effects of recessions? And the answer is, I think, nothing. Because we just used country-level data – that is what we had. There is a clear need to document what it means for health inequities. One would maybe hypothesize that the effect would be an increase in health inequities, because we have the assumption that it's the poor that are hit worst in a recession and hence, that should – other things being equal – lead to growing socioeconomic inequities in health. Now that is not always the case, however, as shown by the very few studies that have looked at the effects of economic crises in the past, using micro-data. So even if you just limit the effect of recession on the distribution of what we consider the socioeconomic determinants of health – take income or education, for example – it is by no means a given that it is always the poor that are hit worst in a recession. It certainly also depends on policy, and how policy responds. And so the post-crisis distribution of income may not necessarily be worse than the pre-crisis one. So again we need to first take this step and then assess that. And that might already take us a step closer to assessing the ultimate impact.* (Marc Suhrcke)

> *The displacement issue [in relation to the ban on smoking in public places]… the other argument I have heard is not simply that people are more exposed in their home to smoking, but if you do reduce smoking you have an effect on diet. So you might actually be contributing to some of the obesity trends. So, as over time smoking has gone down, obesity has gone up – there just might be a causal pathway between those two. So it's another type of displacement that could be going on.* (Jennie Popay)

The health inequities impact of universal policies can be examined in different ways. Some researchers seek to assess impacts across the whole of the health gradient, whereas others focus solely on the experiences of specific disadvantaged groups. Yet others seek to explore the different dimensions of inequity – including gender and ethnicity.

> *Whatever key determinants are identified in WHO* Health 2020 *we cannot forget about, or leave on the sidelines, the determinants of gender, racialization and ethnicity and discrimination, because one of the other trends we are seeing apart from individualism and consumerism is increasing intolerance of either religious diversity or racial or ethnic diversity. And I think - apart from that - gender is a bit of an orphan when it comes to social determinants of health.* (Beth Jackson)

> *We talked about the need not to just look at differential effects by income categories but also a more nuanced analysis looking at gender and race and other elements of socioeconomic categorization.* (Rene Loewenson)

Using tracer groups or conditions

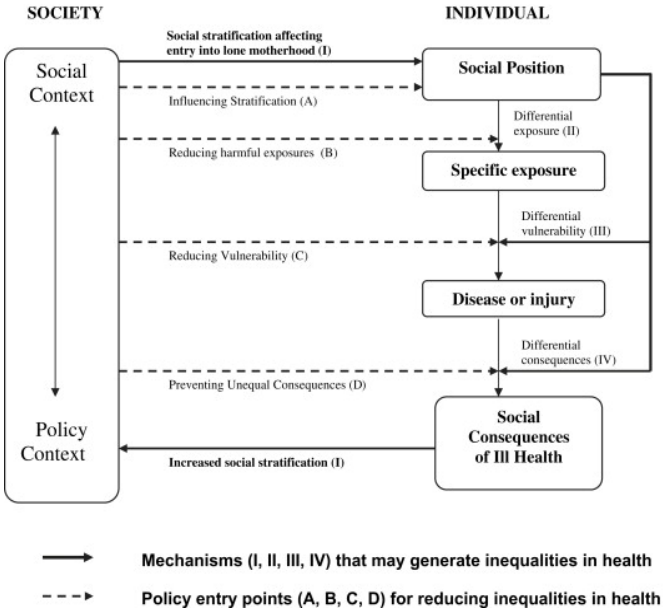
Tracer groups have been used to examine the differential effects of universal policies on the population. The idea is that some groups in the population provide an "early warning" of the impact of changes in universal policies or services, if the groups most reliant on particular policies are carefully selected for study. Lone/single mothers, for example, have been considered as a "litmus test" for family policy components of the social welfare system because they are among the first to be affected by any changes to the system. The Diderichsen model has been used in these tracer studies to elucidate the pathways to inequities between lone/single mothers and those in couples and to identify the most important policy entry points to reduce those inequities (Whitehead, Burström & Diderichsen, 2000; Burström et al., 2010). This requires in-depth analysis of policies that are important to the lives of mothers living in different socioeconomic circumstances, together with analysis of individual-level data on their experiences, health and life chances in contrasting policy contexts (see Box 4.6).

Box 4.6: Studying the impact of universal policies on inequities between lone/single mothers and those in a couple, in Sweden and Britain

The approach we have developed incorporates several components, as described here.

- It follows the experiences of "tracer", or early-warning population groups. In this example, we select lone/single mothers because they have been identified as a good "litmus test" for the workings of the social welfare system – they are among the first to be affected by changes in the system.
- It analyses in depth the policies that are important in the lives of the tracer group (not just the undifferentiated welfare system as a whole). In the case of lone/single mothers, these include family policy, child-care provision, social protection for mothers, family-friendly employment policy, and so on.
- It uses cross-country comparative analysis to capture different policy contexts when the policies being studied are population wide.
- It analyses the relative importance of different mechanisms generating health inequities and policy entry points using the Diderichsen model shown in Fig. 4.6.1.

Fig. 4.6.1 A framework for studying the heath impact of social position and social context and for pinpointing policy entry points



Source: Diderichsen, Evans & Whitehead, 2001.

Our results show that, although the magnitude of the health disadvantage for lone/single mothers compared to those in a couple is similar in the two countries, the pathways leading to that health disadvantage are very different and relate to how different policies operate. Overall, in Britain, about 50% of the health disadvantage of lone/single mothers is accounted for by their greater poverty and unemployment levels (compared to mothers in a couple), which can be traced to the weakness of specific social and labour market policies for this group. In Sweden, the more favourable social policies have protected lone/single mothers from poverty and job insecurity to a much greater extent and these factors only account for between 3% and 13% of the observed health gap, respectively. This finding raises the question: if not poverty and unemployment, then what? Is it "time poverty"? Is it more precarious work? Is it lower access to social support or lone/single mothers compared with those in a couple? These questions lead to further studies.

Sources: Whitehead, Burström & Diderichsen, 2000; Burström et al., 2010.

Section 5: Burning research questions and policies to evaluate as a matter of priority

Participants were asked for their ideas relating to burning research questions and policies that needed evaluating as a matter of priority. The following subsections outline the questions that emerged during the course of the meeting, but there was no opportunity to discuss or put them into order of priority, which leaves the task incomplete.

5.1 What are the differential effects of national policies to deal with economic recession?

The global economic recession and the policy responses to the recession in different countries are likely to have far-reaching impacts on health and health inequities. For instance, in the United Kingdom the response has been to increase austerity measures, leading to cuts in universal services and welfare spending. It is important that the effects of different policy measures are tracked.

> *There are constantly natural experiments happening: we are in the middle of a profound one in our own back yard. And the least we can do over the next six months, a year, two years, five years is begin to document some of the real human stories, of avoidable inequities and suffering that will happen as a result.* (David Stuckler)

5.2 What are the effects of the "Choice" agenda on access to universal services?

In addition to the global economic recession, a further factor contributing to the erosion of universal policies is 21st century consumerism. Consumerist values focus on individualism and consumption, both of which mediate against universal policies. This reinforces policy shifts towards differentiated service provision, with a move increasingly towards more affluent people purchasing services from the private sector, and inadequate public services for the less well-off individuals.

> *We shouldn't lose sight of the predominant 21st century culture, which actually reinforces a move towards choice, consumerism, individualism, general consumption … Attention will also need to be paid to the wider culture that reinforces populations' desires to consume more and more, and to expect everything to be available on tap. And I think that whole consumerist system and*

> *culture is one of the reasons why health systems are potentially also drifting that way.* (Carol Tannahill)

> *There is an increasing thrust for choice and competition, and those drivers may lead to reduced universality of service, or – even if the service remains universal – it might lead to a greater diversity in terms of quality of service provision, and less universal access.* (Carol Tannahill)

Globalization processes may also in future mediate against universal policies at the national level, as policies at the nation-state level give way in many arenas both to more local or regional variation, and more supranational regulation (for example, at the European Union (EU) level). Both of these trends are likely to erode services which have formerly been delivered universally across countries.

> *What do universal policies look like in the global environment, where goods, and services, and trade, and resources, and genetic materials, and biodiversity, and everything are transnational and flowing across borders? Can we continue to design and shape these policies as though they are really fully nationally driven?* (Rene Loewenson)

5.3 What are the features of an equitable health care service?

In the United Kingdom Marmot review of health inequities (Marmot, 2010), access to evidence-based treatments and technologies was highlighted as an important issue, as well as the need to address the wider determinants of health. In universal health care systems, access to new technologies is generally good, but there tends to be a time-lag, with the most advantaged groups accessing new treatments first. In the context of the health care system restructuring that is being undertaken in various countries, a useful research area would be to explore changes in the time to access new technologies by socioeconomic status.

> *[M]aybe we should be thinking about what we would need to do to contemporary health care systems to make them more equitable. And I was thinking about the work that the Department of Health did in England, about what would need to be done to achieve its health equity targets in relation to life expectancy and infant mortality and, although rightly they drew attention to some of the important social determinants of health, namely poverty, they also drew attention to the inequitable access to important, proven health care technology. Before we too readily accept the view that the health care reform either doesn't matter, or it can only be damaging, perhaps we should give more thought to how we need to reform it in the future, to make it more equity focused*. (Ken Judge)

> *There is also a delay in the least advantaged gaining access to the latest technologies. When you look at the access people have, it is the very latest technologies that the most advantaged have greater access to; it then flows through quite quickly, with the least advantaged gaining similar access, but that*

is assuming the existence of a free-at-the-point-of-delivery health care system. If you took that away, the lag-time could be quite great and it might be a valid piece of work to do; a natural experiment, looking at lag-times around the world, comparing access to new technologies between the best- and worst-off. (Peter Goldblatt)

5.4 What are the features of evaluations which have led to both reliable and misleading findings, and what are their effects on policy?

In some instances, similar interventions or policies have been the subject of multiple evaluations – smoking bans in public places are perhaps the best example of this. It may be possible to systematically review these evaluations to explore the contextual factors, and the characteristics of the evaluations themselves, and to relate these to the outcomes of the evaluation. For instance, it may be possible to identify characteristics of evaluations that lead to reliable findings, which are subsequent drivers of policy change.

So that is why it might be useful to find 'killer examples' of really good, incontrovertible evaluations, if there are such things for natural experiments, and also examples where they have been misleading in terms of the findings that they have produced. (Mark Petticrew)

One thing that might be worth considering is to ask ourselves what examples are there of fairly similar natural experiments that have been subjected to multiple evaluations, and systematically reviewing the methods and results arising from those evaluations, to enable us to make more general observations about some of the pitfalls associated with drawing the wrong inferences. From the large number of evaluations associated with the implementation of smoking bans in public places, for instance, it is clear that there is a bit of a dispute now in the literature, about the size of the beneficial effects, and how quickly they can happen. There is also the possibility of the ***displacement effects*** *of smoking bans; one of the fears particularly from* ***an equity perspective*** *is that when you introduce bans on smoking in public places, you actually increase exposure in disadvantaged groups to smoking in their homes. I think it would be a relatively short list of natural experiments in the last few decades, which have been subject to multiple evaluations, and might lend themselves to some kind of systematic review that would be very helpful.* (Ken Judge)

Another example of this relates to Hilary Graham's work on teenage pregnancy (Graham & McDermott, 2005), in which it was possible to explore the contrasting policy implications that resulted, depending on whether one considered the quantitative or qualitative evidence.

What Hilary Graham did was to look at the quantitative evidence, and what the policy implications concerning teenage mothers were in the British policy context.

> *(In Britain there have been many quantitative studies on teenage mothers because it's seen as a problem). But then she identified qualitative studies on teenage mothers, and she found ten really high-quality studies, and she did a meta-ethnography – like a meta-analysis but with the qualitative data – and she showed that the qualitative studies were reporting something completely different and had completely different policy implications. We need more of that sort of nuanced analysis.* (Margaret Whitehead)

5.5 What are the essential elements of an evaluation methodology that capture the equity impact of a universal policy, including the historical context prior to policy implementation and the displacement effects of the implemented policy?

In a systematic way, we need to identify and draw together lessons learnt from past evaluations that have used the various promising methodological approaches, such as the "Resilience" or "Inspector Morse" approaches, exploring how to incorporate context and implementation issues.

5.6 What lessons can we learn from collaborating with (for example) historians and archaeologists to systematically capture how the historical context of a "place" influences the impact (either positive or negative) of an implemented universal policy?

The context in which a policy is implemented is of critical importance, and perhaps in public health we have not sufficiently utilized methods from other disciplines to understand contextual factors.

> *As Jenny has eloquently said on many occasions, the history of place and the history of cultures are such important determinants of the features that we are trying to analyse; I think there is something to be said for a more scientific look at the way we could do that, and learning from the historians and archaeologists and others who piece together things from fragments of facts.* (Mike Kelly)

5.7 What lessons can we learn from economists in assessing the equity impact of a universal policy, factoring in the historical context of the "place"?

An opportunistic tracer group for evaluation – children in low-income households

Using tracer groups that are particularly sensitive to the effects of certain policies – as early indicators of the possible effects of policies on the rest of the population – has already been discussed. Exploring the effects of multiple changes to welfare policies – using young children as tracer groups – would be a suitable area for further research.

> *You could potentially have young children getting a pupil premium attached to them where the housing benefit cap might be affecting their households and then factor in something else about their situation. Then the idea of tracing a group of young children in low-income households which might be affected by a whole range of these changes – some positive, some negative – would be quite an interesting way to approach this, rather than taking just the pupil premium, because then you would get a sense of the net gain (or not).* (Jennie Popay)

5.8 Researching knowledge exchange issues

Further work is also required to understand the "dialogue of the deaf" between policy-makers and the academic community and, as part of this, why politicians fail to translate evidence on universal policies into action, as well as the evidential or advocacy levers required to change this.

> *Ministers … clearly have within their mindset the capacity to think about equality and fairness and social justice, on this broader front, but they seem to lack the capacity to connect this with what they think about those health outcomes.* (Ken Judge)

Section 6: Establishing a longer term partnership

Participants agreed to disseminate and develop the ideas generated in the Liverpool Expert Group Meeting in three key areas, as detailed here.

Participation in developing publications from the meeting. This could include the report of the meeting itself, which was envisaged as capturing the central themes and arguments from the three days of discussions on why and how (and what exactly) to evaluate universal policies for their health equity impact. This would then feed into the development of guidance for WHO Regional Office for Europe purposes on theory and methods for evaluating differential health and other impacts of universal policies. The development process would need to include the identification of gaps in the material generated over the course of the three-day meeting and the commissioning of pieces of work by willing participants to fill those gaps. It was suggested that a journal article summarizing the guidance (as with previous MRC evaluation guidance) would also be useful.

Collaboration on researching "burning questions". Several priority evaluation questions were identified by participants, as outlined in section 5, and many participants were keen to collaborate further on addressing these questions. What is needed now is a process among participants of prioritizing questions for evaluation, as Mark Petticrew suggested (and as reported in section 3): "to focus closely on the 'evaluability' of initiatives and to consider conducting fewer – but more useful – intervention studies". Burning questions identified included the differential impact of policies to tackle economic crises; current health sector reforms in Europe; and regional/city-wide policy impacts. The participants also noted the need to be alert to opportunities to exploit specific natural policy experiments that crop up over the next 12 to 24 months, and to treat members of this expert group as potential collaborators when forming plans for evaluation. Funding for such evaluation was discussed and EU interest in some of the methodological issues was noted. Several participants were already cooperating on a current EU Seventh Framework Programme application to develop methods for evaluation of natural policy experiments for their differential impact, and it was felt that this application was something to build on and develop further.

Feeding into WHO strategy development processes. Participants were keen that the work stimulated by this meeting should inform the development of strategy at WHO European level and beyond. Several opportunities were identified by WHO officers, including feeding into the current WHO European Strategy for Health 2020; informing the WHO European Region Review of Social Determinants and the Health Divide, and interacting with individual countries, for example, by working with them as country 'test beds' for the evaluation guidance being developed. These WHO opportunities were explained as follows:

There will be a European Health 2020 and social determinants of health platform, and that platform will be collaborative, participative and supported by WHO, both globally and within the European Region. It will be supported by the European Review on the Social Determinants of Health and the Health Divide, which Sir Michael Marmot at University College London (UCL) will chair. And then, with the EU, the Organisation for Economic Co-operation and Development (OECD), the World Bank, there will be various advisory fora, some evidence consortia, and some partnership panels. In the context of this meeting these will be some of the key platforms for potential collaboration. And then, we also envisage some country innovation and test sites because we want to work with countries to show that this actually can and does work, and to feed in our own and all of your analytical ideas, but also to learn from the implementation of those ideas in innovative and test sites. (Richard Alderslade)

We could try to build with some Member States structured learning exchanges in the context of cross-government reviews. (Richard Alderslade)

In terms of Health 2020, the WHO European Region Review of Social Determinants and the Health Divide and the New Public Health Strategy for Europe ... we can collectively contribute to this process... certainly we must support governments and policy-makers in their own policy development and review. We could do that perhaps through developing guidelines, in evaluating universal policies. But if you are a minister of health in a country in central Asia and we come along and say this is something we really want you to take seriously, first of all you would have to explain this concept and second you would have to provide some practical assistance for the implementation of it, such as the development of guidelines in evaluating universal policies, and particularly those relating to the social determinants and health inequities ... with some sensitivity to the particular historical developmental, political, and cultural context. (Richard Alderslade)

WHO are trying to incorporate systems thinking into health services and public health. There has been some discussion about the value of case studies of systems thinking that can be used as examples. And there were two sides to the discussion: (1) we don't need more proof of concept, we just need to go and do more of it; and (2) we don't need more proof of concept, but there are some people that still need convincing – there are some people who are on side, but not with their hearts yet; they are not advocates. And to get those people to become advocates, people like us need to be able to point to cases, and say 'Look, it made a difference there', or 'It did something there', or 'Look how useful it was here'. (Alan Shiell)

The group could support WHO in developing materials that explain basic concepts relating to universal policies and the social determinants of health, along with their evaluation. These materials could fulfil an important role in engaging policy-makers in Member States in which these issues are not on the mainstream health agenda. Key elements of this could include an expert discussion paper on the evidence base, and methods for evaluating complex

interventions; identification of candidate policies; and the provision of direct support to governments undertaking such policies.

In addition, suggestions from specific countries included the following points.

- Tone Torgersen suggested that the Norwegian group would be interested taking on an active role in supporting development work on the issue of inequities and the role of universal policies.

  If there is an initiative on these universal strategies of course that would be very interesting for Norway to get involved in ... it may be possible for us to support it, but I guess it would be good for the policy-makers and the director-generals and all those to have a more formal role of this network to take forward into the review. (Tone Poulsson Torgersen)

- Carol Tannahill suggested that her group in Glasgow could play a role in terms of coordinating and connecting other groups representing regions or cities that are interested in undertaking subnational comparisons.

  A few people have said they would be interested in being points of contact for other regions or cities, so if you are interested then let me know and we can even see what we can do in the following up the subnational comparisons. We would be delighted to do that. (Carol Tannahill)

- Maciek Godycki-Cwirko identified an opportunity for dissemination to the European primary care community.

  I keep working on introducing the concept of equity to general practitioners (GPs) at the European level and next year our College is organizing a European congress of GPs and ... a session on equity, especially in the health care systems context ... will be a very (good) opportunity; it will be in Warsaw in September 2011. (Maciek Godycki-Cwirko)

- The group could link to the 'learning networks' that Chris Brown[3] has been developing in partnership with countries, research institutions and policy-based community groups, strengthening the use of evidence in developing interventions to address health inequities.

  It is extremely important that the thinking that has gone on in the last two days of this meeting feeds into the way in which we develop the evidence (and) that it informs the scientific basis for moving forward on the review, and on Health 2020 ... to inform the partnership working that will be needed to develop Health 2020, and the WHO European Region Review of social determinants. It's at that stage of engagement with individual countries

[3] Programme Manager, Governance for Social Determinants of Health/Equity in the WHO European Office for Investment for Health and Development (WHO Venice Office).

working on a case study basis, but in partnership with those countries, that we need to be very clear about the strength of evidence that we are bringing to the table and about the way in which any test sites are then evaluated and how evaluation is carried through. (Peter Goldblatt)

The learning networks that Chris Brown has been supporting across Europe through the case study work could now move a step forward by linking them much more closely with relevant evidence consortia. For example, the discussion we had on tracer populations yesterday might be a really interesting way of connecting in the kind of routine evidence people are using for planning – with a much more specific focus on more scientific evidence around inequities – that then have a much more direct loop back into practice. (Rene Lowenson)

Section 7: Conclusion

The main focus of the meeting was on conceptual and methodological challenges of evaluating universal policies for their impact on social determinants of health and the reduction of health inequities. We considered why this issue was important at this point in time; what the main challenges for evaluation were; what methodological approaches looked promising and could be refined; priorities for evaluation; and how we could establish a longer term partnership to advance the identified evaluation agenda.

The main messages coming out of the meeting include the following points.

- We use the term "universal policies" here to mean ones that are applied to the whole population, but – crucially for our concern with health inequities – they may have differential impact or consequences.

- We need evaluations of universal policies for several important reasons, including those listed here.
  - These policies have the capacity to exert a profound influence on health and health inequities, both positive and negative in nature; precisely because they affect many people, it is possible to detect a significant impact at the population level.
  - Evaluation of such policies for their health equity impact has been relatively neglected, compounding the "inverse evidence law".
  - The advent of the global financial crisis makes the evaluation more urgent, to inform the debate on the most efficient and effective use of scarce resources and to challenge the current erosion of universal policies with evidence.

- Evaluation of universal policies for their health equity impact poses significant challenges, including the following aspects.
  - The defining characteristic of a universal policy – that it is applied across the whole population – precludes a true experimental design in which a comparison is made with an unexposed control group.
  - Time-lags and long causal chains.
  - The difficulty of making links between policy events and outcomes relating to health.
  - The highly complex nature of many universal policies, with multifaceted causal chains that need tracing and unpicking.
  - The importance of the specific context (including organizational context) in which the policies are implemented. This calls for careful account to be taken of context in these evaluations, but also understanding which aspects of findings from the evaluation of universal policy can be transferred to other policy settings.

  - The mismatch between research and policy time frames.
  - The need (and difficulties) of engaging decision-makers from other sectors when dealing with the wider social determinants of health.
  - The tension between "robust" and "good enough" evidence and the need to prioritize interventions for evaluation in the context of scarce resources to conduct research.

- Promising approaches are emerging which have or could be refined for the evaluation of universal policies. These include:
  - innovative ways of exploiting natural policy experiments;
  - using logic models and a systems approach;
  - cross-national comparative policy analyses using policy typologies and tracer groups/conditions;
  - the "Inspector Morse"/"Resilience" approach(es);
  - learning from historical analyses;
  - the value of case studies;
  - drawing insights from Complexity Theory.

- Some priority questions for evaluation include those listed here.
  - What are the differential effects of national policies to deal with economic recession?
  - How are children in low-income households being affected by the current multiple changes to welfare policies?
  - What are the effects of the "Choice" agenda on access to universal services?
  - What are the features of evaluations which have led to both reliable and misleading findings, and what are their effects on policy?
  - What are the essential aspects of an evaluation methodology to capture the equity impact of a universal policy?

The participants agreed to develop ideas generated in the Liverpool Expert Group Meeting in three key ways:

1. participating in developing guidance for WHO purposes on theory and methods for evaluating differential health and other effects of universal policies;
2. collaborating on research into some of the priority questions identified during the meeting;
3. feeding into WHO strategy development processes, including the current WHO Health 2020 European strategy development, informing the European Review of Social Determinants and the Health Divide, and interacting with individual countries to test guidance and learn from their experiences.

Section 8: References and bibliography

Bauld L, Judge K (2002). *Learning from health action zones*. Chichester, Aeneas Press.

Brown SM et al. (1975). Effect on mortality of the 1974 fuel crisis. *Nature*, 257(5524):306–307.

Burström B, Macassa G, Oberg L et al. (2005). Equitable child health interventions: the impact of improved water and sanitation on inequalities in child mortality in Stockholm, 1878 to 1925. *American Journal of Public Health*, 95(2):208–216.

Burström B et al. (2010). Health inequalities between lone and couple mothers and policy under different welfare regimes – the example of Italy, Sweden and Britain. *Social Science and Medicine*, 70:912–924.

Burström B, Öberg L, Smedman L (2011). Policy measures and the survival of foster infants in Stockholm 1878–1925. *European Journal of Public Health*, 13 February: doi: 10.1093/eurpub/ckr002.

Campbell D (1969). Reforms as experiments. *American Psychologist*, 24:409–429.

CSDH (2008). *Closing the gap in a generation: health equity through action on the social determinants of health. Final report of the Commission on Social Determinants of Health*. Geneva, World Health Organization (http://www.who.int/social_determinants/final_report/en, accessed 2 August 2011).

Dahlgren G, Whitehead M (1993). *Tackling inequalities in health: what can we learn from what has been tried?* Working paper prepared for the King's Fund International Seminar on Tackling Inequalities in Health, September, Ditchley Park. London, King's Fund (mimeo).

Dahlgren G, Whitehead M (2007). *European strategies for tackling social inequities in health. Levelling up. Part 2.* Copenhagen, WHO Regional Office for Europe.

Diamond J, Robinson J, eds. (2010). *Natural experiments of history.* London, The Belknap Press of Harvard University Press.

Diderichsen F, Evans T, Whitehead M (2001) The social basis of disparities in health. In: Evans T et al., eds. *Challenging inequalities in health: from ethics to action*. New York, Oxford University Press.

Edelman M (1971). *Politics as symbolic action*. Chicago, Markham Publishing Company.

Edelman M (1977). *Political language: words that succeed and policies that fail.* New York, NY, Academic Press.

European Commission (2009). *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. Solidarity in health: reducing health inequalities in the EU*. Brussels, European Commission (COM (2009)567/4).

Graham H, McDermott E (2005). Qualitative research and the evidence base for policy: insights from studies of teenage mothers in the UK. *Journal of Social Policy*, 35(1):21–37.

Hawe P, Ghali L. Use of social network analysis to map the social relationships of staff and teachers at a school. Health Educ Res. 2008 Feb;23(1):62-9.

Hawe P, Shiell A, Riley T (2009). Theorising interventions as events in systems. *American Journal of Community Psychology*, 43(3–4):267–276.

House of Commons Health Committee (2009). *Health inequalities. Third report of session 2008–09. Volume 1*. London, The Stationery Office Limited (http://www.publications.parliament.uk/pa/cm200809/cmselect/cmhealth/286/286.pdf, accessed 2 August 2011).

Judge K et al. (2005). *Health inequalities: a challenge for Europe. An independent, expert report commissioned by, and published under the auspices of, the UK Presidency of the EU*. London, Central Office of Information (http://eurohealthnet.eu/sites/eurohealthnet.eu/files/publications/pu_2.pdf, accessed 26 July 2011).

Kelly MP et al. (2007). *The social determinants of health: developing an evidence base for political action*. Santiago, WHO Measurement and Evidence Knowledge Network.

Lundberg O et al. (2008a). *The Nordic experience: welfare states and public health (NEWS)*. Stockholm, Stockholm University Centre for Health Equity Studies (CHESS) (Health Equity Studies No. 12).

Lundberg O et al. (2008b). The role of welfare state principles and generosity in social policy programmes for public health: an international comparative study. *Lancet*, 372(9650):1633–1640.

Marmot M (2010). *Fair society, healthy lives: a strategic review of health inequalities in England post-2010*. London, University College London (http://www.marmotreview.org/english-review-of-hi, accessed 10 August 2011).

MRC (2008). *Developing and evaluating complex interventions: new guidance*. London, Medical Research Council.

MRC, Petticrew M (in press). *Using natural experiments to evaluate population health interventions: guidance for users and producers of evidence*. London, Medical Research Council.

Nutbeam D (2003). How does evidence influence public health policy? Tackling health inequalities in England. *Health Promotion Journal of Australia*, 143:154-158.

Petticrew M et al. (2005). Natural experiments: an under-used tool for public health. *Public Health*, 119:751–757.

Popay J, ed. (2006). *Moving beyond effectiveness: methodological issues in the synthesis of diverse sources of evidence.* London, National Institute for Health and Clinical Excellence (NICE).

Popay J, Whitehead M, Hunter D. (2010). Injustice is killing people on a large scale – but what is to be done about it? *Journal of Public Health*, 32:148–149.

Shiell A (2010). Market failure is bad for your health but social injustice is worse. *Journal of Public Health*, 32:12–13.

Shiell A, Hawe P, Gold L (2008). Complex interventions or complex systems? Implications for health economic evaluation. *British Medical Journal*, 336(7656):1281–1283.

Smith RD, Petticrew M (2010). Public health evaluation in the twenty-first century: time to see the wood as well as the trees. *Journal of Public Health*, 32(1):2–7.

Stuckler D et al. (2009). The public health effect of economic crises and alternative policy responses in Europe: an empirical analysis. *Lancet*, 374:315–332.

Whitehead M (2007). A typology of actions to tackle inequalities in health. *Journal of Epidemiology and Community Health*, 61:473–478.

Whitehead M (2010). The right wood, but barking up the wrong tree. *Journal of Public Health*, 32:16–17.

Whitehead M, Burström B, Diderichsen F (2000). Social policies and the pathways to inequalities in health: a comparative analysis of lone mothers in Britain and Sweden. *Social Science and Medicine*, 50:255–270.

Whitehead M et al. (2009). *Helping chronically ill or disabled people into work: what can we learn from international comparative analyses? Final report to the Public Health Research Programme*. London, Department of Health (http://www.york.ac.uk/phrc/PHRC%20C2-06_%20RFR.pdf, accessed 31 July 2011).

Appendix 1: Meeting Agenda

Tuesday, 2 November 2010

Venue: Foresight Centre, University of Liverpool

Session One: Tuesday afternoon, 2 November 2010: 13:30–17:30

Chair: Margaret Whitehead

Focus of the session: Importance of tracking differential impact, challenges and promising approaches

13:30–13:45: Welcome, general remarks, and introduction to the purpose of the meeting and agenda: *Margaret Whitehead*

13:45–14:15: Participants introduce themselves and agree/amend agenda

14:15–14:30: Why is this issue important? The need to track impact of universal policies from a research and strategy perspective: *Margaret Whitehead*

14:30–14:45: Why this issue is important from a policy perspective: *Tone Poulsson Torgersen* and *Richard Alderslade*

14:45–15:00: General discussion to identify key points for rationale

15:00–15:30: Break

15:40–16:00: Evaluation challenges and emerging approaches: "Beyond the MRC guidance on evaluating complex interventions and natural experiments": *Mark Petticrew*

16:00–16:10: Discussion

16:10–16:30: Complex interventions in complex systems: insights from economics and public health: *Alan Shiell*

16:30–16:40: Discussion

16:40–17:00: Challenges of evidence synthesis of complex public health interventions: *Mike Kelly*

17:00–17:20: General discussions to identify key points

17:20–17:30: Summing up of Session One

Wednesday, 3 November 2010

Session Two: Wednesday morning, 3 November 2010: 09:00–12:30

Chair: **Peter Goldblatt**

Focus of the session: Examination of approaches that have been tried and what we can learn from them

9:00–9:10: Introduction to the day

9:10–10:30: Evaluation of national social welfare and economic policies

9:10–9:55: Examples of evaluations and methodological insights on:
- social welfare systems: *Olle Lundberg*
- economic crises: *David Stuckler and Marc Suhrcke*
- using policy entry points and typologies: *Margaret Whitehead*

9:55–10:10: Policy discussants/reflections: *Rene Loewenson*

10:10–10:30: Open discussion on issues of national policy evaluation

10:30–11:00: Break

11:00–12:30: Evaluation of area-based/community initiatives outside health system

11:00–11:45: Examples of evaluations and methodological insights on:
- Health Action Zones: *Ken Judge*
- evaluating community and social interventions aimed at more distal determinants of health: *Alan Shiell*
- evaluating New Deal for Communities and community engagement: *Jennie Popay*

11:45–12:00: Policy discussants/reflections: *Tatjana Buzeti* and *Beth Jackson*

12:00–12:30: Open discussion on issues of evaluation of area-based initiatives

12:30–13:30: Lunch

Continuation of Session Two: Wednesday afternoon, 3 November 2010: 13:30–15:00

Chair: Nicole Valentine/Rene Loewenson

13.30–15.00: Evaluation of impact of health systems

13:30–14:15: Examples of evaluations and methodological insights:
- Health sector reforms in the United Kingdom and the United States: *Mark Exworthy*
- Health sector reforms in Sweden: *Bo Burstrom*
- Equity of access to preventive care in Poland: *Maciek Godycki-Cwirko*

14:15–14:30: Policy discussants/reflections: *Carol Tannahill*

14:30–15:00: Open discussion on evaluation of health system issues

15:00–15:30: Break

Session Three: Wednesday afternoon, 3 November 2010: 15:30–17:30

Chair: Carol Tannahill

Focus of the session: Identification of burning evaluation questions and natural policy experiments/live policies that might help

15:30–16:00: Summing up of insights gleaned from Session Two: *Peter Goldblatt*

16:00–16:30: Most pressing/opportunistic evaluation questions: *Mark Petticrew* and *Jennie Popay* (discussion leaders)

16:30–17:00: Possible natural policy experiments and live policies: *Olle Lundberg and Rene Loewenson* (discussion leaders)

17:00–17:30: Reflections on day and plan for tomorrow: *Margaret Whitehead*

Thursday, 4 November 2010

Session Four: Thursday morning, 4 November 2010: 9:00–12:00

Chair: Richard Alderslade

Focus of the session: Establishing partnerships to take equity evaluation agenda forward

9:00–9:15: Based on the main debates and action areas arising from the meeting, how could these be taken forward by WHO Regional Office for Europe? *Richard Alderslade* and *Peter Goldblatt*

9:15–9:30: Based on the main debates and action areas arising from the meeting, how could these be taken forward by WHO globally? *Nicole Valentine*

9:30–10:00: Open discussion on possible inputs of meeting participants into policy development globally, within the European Region and within specific countries

10:00–10:30: Open discussion on linking to research funding initiatives, both national and cross-national

10:30–11:00: Break

11:00–11:30: Who will do what and when? Next steps: publication

11:30–12:00: Summing up and concluding remarks: *Margaret Whitehead*

Close of meeting

The WHO Regional Office for Europe

The World Health Organization (WHO) is a specialized agency of the United Nations created in 1948 with the primary responsibility for international health matters and public health. The WHO Regional Office for Europe is one of six regional offices throughout the world, each with its own programme geared to the particular health conditions of the countries it serves.

Member States

Albania
Andorra
Armenia
Austria
Azerbaijan
Belarus
Belgium
Bosnia and Herzegovina
Bulgaria
Croatia
Cyprus
Czech Republic
Denmark
Estonia
Finland
France
Georgia
Germany
Greece
Hungary
Iceland
Ireland
Israel
Italy
Kazakhstan
Kyrgyzstan
Latvia
Lithuania
Luxembourg
Malta
Monaco
Montenegro
Netherlands
Norway
Poland
Portugal
Republic of Moldova
Romania
Russian Federation
San Marino
Serbia
Slovakia
Slovenia
Spain
Sweden
Switzerland
Tajikistan
The former Yugoslav Republic of Macedonia
Turkey
Turkmenistan
Ukraine
United Kingdom
Uzbekistan

WHOLIS Number: E95912

How can the health equity impact of universal policies be evaluated?

World Health Organization
Regional Office for Europe

Scherfigsvej 8, DK-2100 Copenhagen Ø, Denmark
Tel.: +45 39 17 17 17. Fax: +45 39 17 18 18. E-mail: contact@euro.who.int
Web site: www.euro.who.int