WORLD HEALTH ORGANISATION

COMMISSION ON SOCIAL DETERMINANTS OF HEALTH

MEASUREMENT AND EVIDENCE KNOWLEDGE NETWORK

THE DEVELOPMENT OF THE EVIDENCE BASE ABOUT THE SOCIAL DETERMINANTS OF HEALTH

Michael P Kelly, Josiane Bonnefoy, Antony Morgan, Francisca Florenzano

The National Institute for Health and Clinical Excellence (NICE) (UK) and the Universidad del Desarrollo (UDD) (Chile)

May 2006
THE DEVELOPMENT OF THE EVIDENCE BASE ABOUT THE SOCIAL DETERMINANTS OF HEALTH

May 2006
1. Introduction

1.1. Commission on the Social Determinants of Health
In 2005, the Director General of the WHO set up a global Commission on the Social Determinants of Health (CSDH). The objective of the Commission was to achieve policy change by learning from existing knowledge about the social determinants of health (SDH) and turning that learning into global and national political and economic action. To facilitate the learning a number of Knowledge Networks (KNs) were established by WHO to synthesize knowledge about the social determinants of health. This paper has been prepared by one of those knowledge networks viz. the Measurement and Evidence Knowledge Network. The purpose of this paper is to articulate a series of methodological, theoretical and epistemological principles that will help to inform the development of the evidence base about the social determinants of health which all the Knowledge Networks will be working to construct.

1.2 Initial Conceptual Ideas
Solar and Irwin (2005) developed a discussion paper for the CSDH ‘Towards a conceptual framework for analysis and action on the social determinants of health’ to set out the conceptual foundations for the work of the Commission. It put forward a schema (drawn from existing models and frameworks) for the social determinants of health which aimed to:

- clarify the mechanisms by which social determinants generate health inequities
- show how the major determinants relate to each other
- provide a framework for evaluating which social determinants of health are the most important to address
- map specific levels of intervention and policy entry points for action.

The framework also highlighted three key issues that need to be addressed if effective action is to be taken on the social determinants of health:

1. to distinguish between the structural (e.g. income and education) and intermediate (e.g. living and working conditions, population behaviour, food availability) determinants of health;
2. to understand and make explicit what is meant by the socio-political context (encompassing a broad set of structural, cultural and functional aspects of the social system which impacts on individuals);

3. to take account of the actions that need to be taken at different levels (macro, meso, micro) in order that inequalities in health can be tackled (i.e. to alter the configuration of underlying social stratification, and those policies and interventions that target intermediate health determinants).

The present paper considers a number of intellectual principles which need to be articulated in order to tackle the issues identified by Solar and Irwin.

2. Principles

Principle 1: A commitment to the value of equity

Globally there have been impressive improvements in overall indicators of health over recent decades. None the less, health inequalities within and between countries persist and in many cases have widened and continue to widen (WHO, 2004). This is in spite of the fact that the pursuit of equity and the reduction in health inequalities has been a goal of some national (Graham, 2004a; 2004b) and a number of international policies (WHO, 1981; 1985; 1998a; Ritsatakis, 2000; Braveman et al., 1996; Braveman, 1998; United Nations, 2000) over the years.

The first principle for the development of the evidence base for the social determinants of health is a statement of the value position of equity against this background. The explicit value is that the health inequity that exists within and between societies is unfair and unjust. This is not a scientifically or rationally derived principle; it is a political position which asserts the rights to good health of the population at large and to the equitable disbursement of the benefits of social and medical advances. It stands in contrast to the value position that argues that differences in health are a consequence (albeit perhaps an unfortunate consequence) of the beneficial effects of the maximisation of individual utility in the market. It is important to note therefore at the outset, that individual and collective utilities may be at odds politically, with the equitable right to health.

The position taken here is that systematically differential patterns of health outcomes which have their origins in social factors are unfair and unjust and the social factors
which cause this state of affairs are also unfair and unjust. The explicit value position is that this is morally indefensible and that there is an imperative to find solutions. Furthermore, because these factors are social and they are the product of human agency they are also potentially changeable through human agency.

However, this is a difficult task. Discussion about social determinants of health takes place within a sometimes explicit, but usually implicit, tension between the competing claims of rights and utilities mentioned above. The claims and counter claims of rights and utilities are irresolvable through rational discourse because each appeals for justification to grounds which the other rejects, although in their own terms both positions are logically consistent. In short, to uphold one person’s or group’s rights is to interfere with some other individual or group’s utilities – and vice versa (Macintyre, 1984). To assert the right to health of a group in society who experience poor health by virtue of social arrangements, is not only to appeal to an argument about social justice, but also to make the case for taking something away from other individuals or other social groups who enjoy good health as a consequence of those same social arrangements. This might be in the form of taxation, or some other infringement of, or restraint on economic liberty. In other words the rights to health, and the right to the equitable distribution to health come at a cost borne usually by others. The political position of this paper, which has other epistemological consequences, is that it is a cost worth paying. This is of course a position which will not find universal acceptance, because the principles of social and moral justice on which it is based are not givens; they are political assertions.

There are two implications which flow from this. First, to aim to eradicate inequity in health logically means also eliminating the social determinants of health inequalities. These determinants are the various dimensions of social disadvantage and difference (Braveman, 2003) - income, employment, education, ethnicity, gender, housing, and environment for example (Graham, 2000; Marmot & Wilkinson, 1999; Shaw et al., 1999; Solar & Irwin, 2005 for a review). Therefore a very precise means of describing the dimensions of social difference and disadvantage is required. Second, the language used here is one of cause. A very precise way of describing the causal pathways to health inequity is also required. These two implications are considered below.

The term inequity is preferred to inequality in this paper. The words will be defined following Whitehead (1992). Whitehead describes health inequality as ‘measurable
differences in health experience and health outcomes between different population groups – according to socioeconomic status, geographical area, age, disability, gender or ethnic group’. Inequality is about objective differences between groups and individuals measurable by mortality and morbidity. Whitehead defines ‘health inequity’ as differences in opportunity for different population groups which result in for example, unequal life chances, access to health services, nutritious food, adequate housing etc. These differences may be measurable; they are also judged to be unfair and unjust (Whitehead, 1992). It is noted that the terms health inequalities and health inequities are used in different ways in different societies and by different authors (Leon et al., 2001). However, for the present purposes, Whitehead’s distinction will serve usefully.

Principle 2: Taking an evidence based approach

The second principle is a commitment to an evidence based approach. It is taken that is axiomatic that an evidence based approach offers the best hope of tackling the inequities that arise as a consequence of the operation of the social determinants. Further it is assumed that the evidence will provide the basis for understanding and the basis for action (Greenhalgh, 2001).

There are however a number of challenges. There is a very rich literature describing health inequalities and the social determinants of health, especially in developed countries (Graham, 2000; Marmot & Wilkinson, 1999; Shaw et al., 1999; Solar & Irwin, 2005 for a review). However there is a dearth of good scientific studies explaining what can be done to reduce health inequalities (Millward et al., 2003). There is a lack of systematic studies of the effects of policy on inequity. The contours of inequality and social difference and disadvantage are not well described. The degree to which changes in inequalities can be measured is ill defined (Killoran & Kelly, 2004). The difference between the determinants of health and the determinants of inequalities in health is often confused (Graham & Kelly, 2004; Graham, 2004a; 2004b; 2004c). The health of populations and the health of individuals is frequently elided (Heller, 2005). And, finally, the links between the proximal, intermediate and distal determinants of health are poorly conceptualized and integrated into research (WHO, 2004).

In the face of these difficulties an evidence based approach means finding the best possible evidence about the social determinants (NHMRC, 1999). The most sophisticated and technologically advanced search strategies and systematic review
procedures should be used (Glasziou et al., 2004; Jackson & Waters 2005a; 2005b) along with traditional forms of scholarship. The definition of best evidence should be made on the basis of its fitness for purpose and on the basis of its connectedness to research questions (Glasziou et al., 2004). Those research questions are the ones which deal with the effectiveness of interventions to change the social determinants.

There are of course some important caveats about the evidence based approach. There will be gaps in this evidence and some parts of it will be more powerful than other parts. There will have to be a recognition that strength of evidence alone is not sufficient as a basis for making policy (NHMRC, 1999) and that it is possible to have very good evidence about unimportant problems and limited or poor evidence about very important ones. Therefore a distinction must be drawn between absence of evidence, of poor evidence and evidence of ineffectiveness. The two former are not the same as the latter. It will need to be recognised that the links between scientific knowledge and policy and practice are not linear and that the scientific evidence base is generally imperfect in its own methodological, theoretical and empirical terms. Consequently the connection between evidence and policy and practice inevitably involves matters of judgements (Kelly et al., 2004). Therefore the strength of evidence alone should not drive the strength of policy recommendation (Harbour & Miller, 2001). Linking evidence based to health policy will require the identification of appropriate and culturally sensitive mechanisms (Rawlins, 2005; Briss, 2005). This leads to a commitment to the principle that the application of research findings to non research settings requires an understanding of the local context and the tacit knowledge and the life worlds of practitioners and end users. It also means that evidence hierarchies must be used flexibly.

**Principle 3: Methods and epistemology**
The data and evidence which relate to social determinants of health come from a variety of disciplinary backgrounds and methodological traditions. The evidence about the social determinants comprises a range of ways of knowing about the biological, psychological, social, economic and material worlds. The disciplinary differences arise because social history, economics, social policy, anthropology, politics, development studies, psychology, sociology, environmental science and epidemiology, as well as biology and medicine may all make contributions. However, each of these has its own disciplinary paradigms, arenas of debate, agreed canons, and particular epistemological positions. Some of the contributions of these
disciplines are highly political in tone and intent. And in spite of a great deal of research endeavour and comment as well as practical attempts at problem solution, there is also a great deal which is not known about the causes of inequalities in health.

In short, although the empirical subject matter of the social determinants of health is diverse, that diversity is given an added layer of complexity by the disciplines involved and that those disciplines do not reach an easy consensus on the nature of knowing the material nor its interpretation. When the ways of knowing and understanding within the worlds of policy makers, politicians, NGOs, as well as of the people whose lives are directly affected by the social determinants (Lomas et al., 2005), the degree of complexity could be potentially debilitating. As an evidence base therefore it has a number of problems: it is drawn from a diversity of disciplines using different methods, it is incomplete, and it is biased in various ways, including political and ideological bias. This does not mean it is unusable; it means we must devise ways of sorting out the disciplinary differences, of filling the gaps and of articulating the bias while valuing the diversity.

It is therefore inappropriate to rule out evidence and data a priori on the basis of their disciplinary and methodological provenance. The immediate task is to find the best evidence, from whatever source it comes, defined by the extent to which it has used an appropriate method to answer the research question. It is axiomatic that to assert the superiority of one type of knowing over another will be unhelpful. A range of types of knowledge and knowing will be important (Kelly, 2004; Berwick, 2005). A pluralistic approach will therefore be necessary. A consideration needs to be made about whether what we know is suitable for what we need to do?

The solution is straightforward and has been a premise of western philosophical thought for millennia (Plato, 1974). Humans use different forms of knowing and different forms of knowledge for different purposes. There is no necessary hierarchy of knowledge involved until we need to discriminate on the basis of fitness for purpose. It is necessary to describe the criteria for acceptability and fitness for purpose. The task will involve doing this across a range of different knowledge types. This does not mean that all knowledge and knowing in general, or of the social determinants of health in particular, is of equal value. It means we have to develop multiple criteria to determine fitness for purpose, to judge thresholds of acceptability and critically appraise the knowledge on this basis.
Principle 3 therefore promotes the use of a wide range of methodologies to assess the success of interventions and policies which aim to address the social determinants of health.

Particular attention will need to be paid to the role of qualitative research in assessing the effectiveness of approaches to address the social determinants of health. Two different models to describe the ways in which qualitative evidence contributes to the evidence base for policymaking (Popay, 2005).

1) The *enhancement model* assumes that qualitative research adds something ‘extra’ to the findings of quantitative research – by generating hypotheses to be tested, by helping to construct more sophisticated measures of social phenomena, and by explaining unexpected findings generated by quantitative research.

2) The *epistemological model* views qualitative evidence as making an equal and parallel contribution to the evidence base through: (a) focusing on questions that other approaches cannot reach; (b) increasing understanding by adding conceptual and theoretical depth to knowledge; and (c) shifting the balance of power between researchers and the researched (Popay, 2005 unpublished). Importantly, the epistemological model views qualitative evidence as not necessarily complementing quantitative evidence, but sometimes conflicting with it.

Qualitative research can play two key roles as part of the evidence base for the social determinants of health: (a) providing insights into the subjectively perceived needs of the people who are to be the targets of the interventions and programmes aimed at addressing the social determinants of health and health inequalities (giving people a ‘voice’); and (b) helping to unpick the ‘black box’ of interventions and programmes to deepen understanding about factors shaping implementation, and hence, impact (Roen *et al.*, 2005; Arai *et al.*, 2005). One major difference between the qualitative and quantitative traditions concerns the notion of replicability and generalizability. Obviously generalizability within the qualitative tradition is of a different kind to that which is possible in an experiment or a survey (Popay, 2005). With regard to judging the *external validity* of qualitative evidence, Popay notes: ‘[t]he aim [in the qualitative tradition] is to identify findings which are logically generalizable rather than probabilistically so’ (Popay *et al.*, 1998). It should also be noted that there is a rapidly growing literature on methods for the synthesis of qualitative research.
and of mixed methods research (see for example, Dixon Woods et al., 2004; Popay & Roen, 2003).

There must therefore be a commitment to methodological pluralism and epistemological variability and a commitment to the view that epistemological positions should not be viewed as mutually incompatible. The argument that there is an inherent incompatibility between objectivist and subjectivist approaches is to be explicitly rejected in favour of the view that there are different ways of knowing, and that different ways of knowing can and do play different roles in the ways that human actors use knowledge and information. However, in certain circumstances and for certain purposes some forms of knowing are more practically useful. The polarization of knowledge into objectivist and subjectivist approaches is unhelpful and misleading (See Gomm & Davies, 2000; and Gomm et al., 2000 for a review of helpful ways to describe different methodological approaches). The view that all knowledge is relative and of equal value is to be rejected in favour of a view which defines the relevance and the salience of knowledge, according to its practical value in given circumstances.

**Principle 4: Gradients and gaps**

There are conventionally three different ways in which the inequalities are described: health disadvantage, health gaps and health gradients (for a full discussion of this see Graham, 2004a; 2004b; 2005; and Graham & Kelly, 2004). Health disadvantage simply focuses on differences, acknowledging that there are differences between distinct segments of the population, or between societies. The health gaps approach focuses on the differences between the worst off and everybody else, often assuming that those who are not the worst off enjoy uniformly good health. The health gradient approach relates to the health differences across the whole spectrum of the population acknowledging a systematically patterned gradient in health inequalities.

Conceptually, narrowing health gaps means raising the health of the poorest, fastest. It requires both improving the health of the poorest and doing so at a rate which outstrips that of the wider population. It is an important policy goal. It focuses attention on the fact that overall gains in health have been at the cost of persisting and widening inequalities between socioeconomic groups and areas. It facilitates target setting. It provides clear criteria for monitoring and evaluation. An effective policy is one which achieves both an absolute and a relative improvement in the
health of the poorest groups (or in their social conditions and in the prevalence of risk factors).

However, focusing on health gaps can limit the policy vision. This is why the health gradient is also important. The penalties of inequalities in health affect the whole social hierarchy and usually increase from the bottom to the top. Thus, if policies only address those at the bottom of the social hierarchy, inequalities in health will still exist and it will also mean that the social determinants still exert their malign influence. The approach to be adopted should involve a consideration of the whole gradient in health inequalities rather than only focusing on the health of the most disadvantaged. The significant caveat is that where the health gap is both very large and the population numbers in the extreme circumstances is high, a process of prioritising action by beginning with the most disadvantaged would be the immediate concern.

This approach is in line with international health policy. The founding principle of the WHO was that the enjoyment of the highest attainable standard of health is a fundamental human right, and should be within reach of all ‘without distinction for race, religion, political belief, economic or social condition’ (WHO, 1948). As this implies, the standards of health enjoyed by the best-off should be attainable by all. The principle is that the effects of policies to tackle health inequalities must therefore extend beyond those in the poorest circumstances and the poorest health. Assuming that health and living standards for those at the top of the socioeconomic hierarchy continue to improve, an effective policy is one that meets two criteria. It is associated with (a) improvements in health (or a positive change in its underlying determinants) for all socioeconomic groups up to the highest socioeconomic group and (b) a rate of improvement which increases at each step down the socioeconomic ladder. In other words, a differential rate of improvement is required: greatest for the poorest groups, with the rate of gain progressively decreasing for higher socioeconomic groups. It locates the causes of health inequality, not in the disadvantaged circumstances and health-damaging behaviours of the poorest groups, but in the systematic differences in life chances, living standards and lifestyles associated with people’s unequal positions in the socioeconomic hierarchy (Graham & Kelly, 2004).

**Principle 5: Causes: determinants and outcomes**

Principle 5 is that a basis for developing the evidence is a causal model which crosses from the social to the biological.
What is generally missing in the analysis of social factors and health is the kind of underlying certainty about effectiveness and cause which we have come to expect with respect to clinical medicine. We see instead associational or probabilistic types of explanations (Link & Phelan, 2005; Mechanic et al., 2005). Clinical medicine has its own uncertainties of course. Aetiology is sometimes unknown, tenuous, partial and multifaceted. The effects of treatments are also uncertain (Chalmers, 2004). The disease categories used by medicine to describe pathology, are not essentialist but are nominalist and therefore change and evolve over time. Data and evidence are surrounded by uncertainty (Griffiths et al., 2005), and the skill of the doctor is in the end about working through and with these uncertainties, not resolving them.

But not with standing the uncertain and contingent nature of the understanding of biomedical processes, medicine operates very successfully with an underlying epistemological principle which is that health outcomes have preceding causes and that the isolation of cause is the basis of effective intervention. In the case of inequalities in health, real pathological changes in the human body occur, but in highly patterned ways in whole populations. The assumption made here is that both the pathologies and their patterning have causes. There will be social and biological causes working in tandem. The task is to map that process as a way of developing an explanation. In classic scientific terms there ought to be covering scientific social and biological laws (Hempel, 1965). What needs to be explained is why the biological systems in the human body change in ways that are determined by social as well as biological/biochemical processes. At the heart of the intellectual challenge of the social determination of health and the corresponding inequalities in health is this. The molecules in the human body behave differently according to the social position someone occupies, according to their job, according to their experience of class, gender and ethnic relations, according to their education, and according to a whole range of social factors which impact on them over their life course. Their genetic structure and their immunity, their nutritional status, their resilience, their ability to cope, all act as mediating and biological causal factors, but ultimately there is a plausible causal pathway from a number of social factors or social determinants to biological structures in the individual human body.

In the clinical realm (in which the social is to a significant degree controlled out of the equation) the randomized controlled trial provides the best way of determining what the mechanisms of cause are and what precisely it is, that is effective (Chalmers, 1998). The randomized controlled trial provides the most secure basis for valid
causal inferences about the effects of treatments (Chalmers, 1998). *Inter alia*, to what extent can similar methods be applied in the social realm? The answer to this is that in principle it can, but in practice the available evidence will be very limited. This is because the causal pathways are still to be defined with the appropriate degree of exactness and the covering laws are not known at all. We may have many decades of research ahead of us before the covering laws might be described. However, the challenge of conjoining the social and the biological, and of developing plausible explanatory models demand our immediate attention.

It has been argued that before 1948 clinical medicine was dominated by what today we would call theories and political positions (Cochrane, 1972; Doll, 1998). It is suggested that these theories were tested empirically by individual clinicians, but were never subject to the kind of deep rigorous scrutiny which the clinical trial permits (Greenhalgh, 2001). Effectiveness was in much more tenuous territory than it is today. Doll has argued (Doll, 1998) that 1948 was a watershed because it was the year that the streptomycin trial for treating pulmonary tuberculosis reported. The methodological breakthrough was that effectiveness could be plainly demonstrated. Although of course in 1948, the clinical trial still had many years to go before it found general acceptance (Cochrane, 1972; Egger *et al*., 2001), the fundamental principle was established and causal explanation was in grasp.

The question is to what extent is the study of the social determinants of public health governed by untested theories? To what extent is the study of the social determinants led by theory rather than by evidence of effectiveness? The answer is that it is to a significant degree. This does not render the resulting knowledge worthless: far from it. It does however mean that it is more difficult to interpret and that cause in the sense used here difficult to discern.

In the absence of covering laws the approach to be adopted is that of separating necessary and sufficient conditions and proximal, intermediate and distal causes. The necessary condition is/are the preceding phenomenon which needs to be identified and be described without which the succeeding phenomena will not occur. The sufficient conditions will describe the degree or volume which is required to produce an effect. A true causal model would permit the statement, ‘if A then always B’. By identifying the necessary and sufficient conditions it is possible to develop such statements (Davidson, 1967). A true causal model would also account for the nature of the relationship between A and B. This is what Hempel (1965) called the
covering law. Of course, because the subject matter is going to be surrounded by varying degrees of uncertainty, the initial models or model will be weaker than a true causal model. However, it is the degree of precision of the true causal model that should be the goal, and unravelling necessary and sufficient conditions is the starting point.

With respect to the social determinants of health, we are able to identify some of what are the necessary and the sufficient conditions but the nature of which are which and under what circumstances, is very unclear. The core candidates can be listed relatively easily because the extant literature has explored them at length: occupational exposure to hazards, occupational experience of relations at work (degree of self direction for example), the biological aging process, the experience of gender relations, the experience of ethnic relations including direct experience of racism, home circumstances, degree and ability to exert self efficacy especially through disposable income, dietary intake, habitual behaviours relating to food, alcohol, tobacco and exercise, position now and in the past in the life course, schooling, marital status and socio economic status. These are the media through which the social world impacts directly on the life experiences and exert direct effects on the human body. They in turn are linked to macro variables like the class system, the housing stock, the education system, the operation of markets in goods and labour and so on (see Solar & Irwin, 2005).

However, just listing the factors, neither tells you what the linkages actually are, nor what the covering law is, nor what the biologically plausible relations actually consists of. As Smith (2004) has argued, if we combine all the dimensions of social differences into one construct, like socio economic group, this precludes discussion of the policy relevant options (Smith, 2004), but it also precludes proper explanation. There is clearly an urgent need for these relationships to be modelled.

The problem of multi faceted causation will need to be considered in the modelling process. It is clear that there are likely to be a range of factors involved in the explanation and the component parts of the model will need to be delineated. However, this must not degenerate into simply asserting that it is very complex, because this is no explanation at all (Cohen, 1951). Modelling in a multi factorial way allows the delineation of the necessary and sufficient conditions.
The evil causes evil fallacy (Cohen, 1951) also needs to be avoided in the modelling. Antonovsky (1985; 1987) called this the pathogenic approach. By this he meant, a search for system dysfunction, or the identification of the breakdown of idealized social systems. He argued that the social and medical sciences were dominated by a pathogenic orientation. Applied to health inequity, a pathogenic argument is that health inequities are a pathological deviation from an idealized better state caused by pathogenic mechanisms. The pathogens are usually said to be things like global capitalism, political decisions, failing health care systems and poverty. This is unhelpful on two counts. First, idealised perfect non pathological social systems do not exist, and the pathology which is identified as the cause is not an explanation, it is a political statement about values. The value system is used as the explanation. Now, that it is not to say that the tackling of health inequalities and the associated suffering and premature mortality are not worthwhile things to do, nor that is a value position of which to be diffident (see above). Quite the contrary, it is a prime value which should drive forward research and action. But a value, which determines that something is bad (or good), is not the same thing as an explanation. Nor is it to suggest that the operation of global capitalism, poverty, failing health systems are not part of the causal explanation. But for these factors to be integrated into the explanation they need to be stripped of their ideological content and their roles as necessary and sufficient conditions in a web of causation needs to articulated.

To understand health inequalities we must turn to a concept of cause which has its origins in rationalist thought and which in effect mirrors the kinds of precision about cause which clinical medicine is capable of delivering. This, it will be argued, requires a classical scientific explanation: neither an historical nor a sociological explanation will do (Danto, 1968). This is because the phenomena being explained are not historical or social: they are physical. An explanation which stops at the social level is insufficient for these purposes. We need a model of cause which traverses a number of levels of analysis which academic disciplines traditionally keep separate. Some of the observed patterns which are manifested in mortality and morbidity data are no doubt accounted for genetically or by other biological mechanisms, but it seems inconceivable that the health variations which follow so closely sets of social arrangements could all be accounted for in this way. Other processes are at work and they are amenable to causal analysis involving a pathway from the social to the biological. This does not undermine any other form of analysis like a sociological one which operates at the level of the social, nor does it preclude bringing aspects of the sociological explanations into play: far from it. But the principles of cause should be
applied to the issue in question across the social and biological. In this sense the concern is not really inequities in health, but much more specifically the social determinants of inequities in illness. The research question is to find out what the social determinants of mortality and morbidity are and to describe how they work and how they interact with the biological.

This will lead further to use the distinction between the determinants of health and the determinants of inequalities in health. The commitment to addressing the social determinants of health is often summed up in the phrase ‘tackling the determinants of health and health inequalities’. Such phrases can create the impression that policies aimed at tackling the determinants of health are also and automatically tackling the determinants of health inequalities. What is obscured is that tackling the determinants of health inequalities is about tackling the unequal distribution of health determinants.

Focusing on the unequal distribution of determinants is important for thinking about policy. This is because policies that have achieved overall improvements in key determinants such as living standards and smoking have not reduced inequalities in these major influences on health. As noted above positive trends in health determinants can go hand-in-hand with widening inequalities in their social distribution. As these examples suggest, distinguishing between the overall level and the social distribution of health determinants is essential for policy development. When health equity is the goal, the priority of a determinants-oriented strategy is to reduce inequalities in the major influences on people’s health. Tackling inequalities in social position is likely to be at the heart of such a strategy. It is the pivotal point in the causal chain linking broad (‘wider’) determinants to the noxious agents that directly damage people’s health.

Therefore the model of cause needs to be articulated. The evidence should be interrogated to determine what phenomena are attributed to other phenomena. Are necessary and sufficient conditions specified, is the causal chain concerned with proximal, intermediate or distal causes, and what are the plausibility levels of the proposed mechanisms? In brief are we able to find patterns which point to strong causal or associational relationships? To what degree are we able to discern a consistent direction in the evidence, and to what degree are the patterns of the results or the conclusions of studies broadly similar? Is there a relationship which suggests that more of the exposure or the intervention produces more of an effect?
If there is then we have a much clearer sense of potential cause and are able to map out what the proposed mechanisms are.

The level, or levels of analysis, needs to be identified (Kelly, Charlton and Hanlon, 1993). This means examining the evidence, and regardless of its disciplinary provenance, assessing whether the dynamics of what is described could plausibly work at a physical, societal, organisational, community or individual level. In other words, to what degree is the policy or intervention based on biological, social, technical plausibility? To what extent is it possible to ascertain time periods and the chronology in the evidence? Are the purported relationships logically possible in chronological terms? Do certain events precede others? What dynamic processes in terms of the component parts of social systems are described? This is particularly important in multi factorial explanations, where the sequencing of events may be hidden, or at least difficult to discern and where, as we noted above, multi factorial explanations are often no explanations at all (cf. Brownson et al., 2003).

Principle 6: Social structure

Principle 6 aims to make more explicit the range of dimensions of inequalities that need to be considered when building an evidence base on how best to address the social determinants of health, including ethnicity, gender, sexuality, age, area, community and religion (Anthias, 1990; 1992). These represent linked but separate dimensions of inequality. Whilst these discrete dimensions of social difference are seldom denied as important, they are underdeveloped empirically and theoretically in the literature on social determinants. Consequently, the relationships between the different dimensions of inequality and the ways they interact with each other to produce health effects, are hardly to be found in the extant evidence at all (Graham & Kelly, 2004). This is a point of very considerable importance because, it is clear from the evidence that does exist, that different segments of the population respond very differently to identical public health interventions. This means that we need to anticipate a wide range of responses to policies across and within societies, by virtue of the nature of the variation in populations.

What these different and variable axes of differentiation have in common is that they result in differences in life chances. These differences in life chances are literal: there are marked social differences in the chances of living a healthy life. This has been most systematically captured in occupation-based measures of socioeconomic position – but differences in people’s health experiences and their patterns of mortality are observed.
across other axes of social differentiation. It is an important challenge to develop measures of inequality that embrace these axes. If, as the evidence suggests, dimensions of disadvantage interlock and take a cumulative toll on health, these dimensions need to be summed in order both to map and to understand the health penalty of social inequality.

One of the key principles therefore is to acknowledge and to identify the different axes of social difference (Graham & Kelly, 2004) and that these dimensions overlap (Davey-Smith et al., 2000). Within different axes of differentiation, like gender, different aspects interplay as well, like income access to power and prestige (Bartley et al., 2000). The specific health impacts will be mediated by proximal factors like social position, specific exposures, the nature of specific illnesses and injuries and their social significance in different cultural contexts (Whitehead et al., 2000). The model which will be developed will also need to account for the fact that these different aspects of social difference vary independently of each other. But they also coalesce together in varying ways to produce overall patterns of advantage and disadvantage.

Material and environmental disadvantage accumulate through the life course and in particular childhood disadvantage is associated with disadvantage in later life (Benzeval et al., 2000). The two building blocks which will be used to develop these ideas are those of the life course and the life world. Life course epidemiology shows how socially patterned exposures during childhood, adolescence and early adult life may operate via chains of social, biological or psychological risk (Kuh et al., 2003; Graham & Power, 2004). The purpose of life course epidemiology is to build and test theoretical models that postulate pathways linking exposures across the life course to later life health outcomes (Kuh et al., 2003). The life world is a social space, part physical, but predominantly cognitive and subjective. The life world is where we experience the social structure first hand in the form of opportunities, barriers, difficulties and disadvantages. Schutz (1964; 1967; 1970) conceptualized the totality of the experience of the life world as a series of concentric circles. The innermost circle is the one where the everyday contacts and routines are highly predictable and are therefore taken for granted, which are salient and immediate and which tend most of the time to be the most important. There are more distant parts of the life world. It is important to note that the innermost circle of the life world may not be, and Schutz never suggested it would be, a place that was benign and cosy. It may be violent and bullying. It may be cold and unforgiving. It may be unpleasant and
chronically difficult. It will be the place where discrimination and disadvantage are experienced. However, it constitutes the centre of the existence of the person. This is because life worlds are the building blocks of social life. It is the point where social structure impacts on the individual. The life world is where the causal mechanisms of health inequalities operate, and the pathways to ill health can be described. It is the bridge between the social and the biological.

The importance of this idea for patterning of health is that health is an outcome of the accumulated effects of a variety of social and biological factors which impact on people at distinct periods in the life course in the life world. These factors act both positively and negatively as well as cumulatively.

So what is the model of social structure, if any, in the evidence? This means considering the extent to which the evidence is sensitive to the relations between groups and individuals and in particular the social variations and differences in the population. The important differences along the dimensions of age, gender, religion, caste, occupation, mobility, place, residence, status grouping, and class membership.

Principle 7: Social dynamics
Principle 7 highlights that the social systems and sub systems which make up societies are not static objects, they are constantly changing and therefore the relationships which give rise to the outcome of health inequalities and differences are themselves also changing in terms of their force and in terms of their salience at any given moment.

In compiling evidence across knowledge networks it is important therefore to reflect on the historical dimension with respect to the social determination of health. Whilst inequalities in health seem to be a characteristic of all modern contemporary societies, the shape they currently have is not a given, is not set in stone. It is instead something which changes. The question is whether there is any discernible historical patterning which would help us to understand what is going on, and what the processes involved actually are.

The starting point for such analysis is the path breaking work of Antonovsky (1967). In what was one of the very earliest attempts to review historical and contemporary evidence about inequalities in health in a systematic way, Antonovsky showed that
inequalities were a common feature of all advanced social systems. Examining data from more than thirty international studies he argued that the inescapable conclusion was that social class influenced a person’s chance of staying alive. Historically he noted a variation of about 2:1 between the extremities of the social classes, although he saw this differential narrowing in the mid nineteen sixties. This class differential held even though overall death rates were declining. He noted that whatever the index used, or however the class system was represented, almost invariably the lowest social classes had the highest mortality rates (Antonovsky, 1967).

He went on to demonstrate that there was an important characteristic in the historical differences between the most and the least advantaged across different societies. He observed that where the overall rates of mortality were high, the differences in mortality between the best and the worse off tended to be relatively small. This, he claimed, characterised societies in the early period of industrialization. As rates of economic growth increased, and particularly as industrialization evolved, the patterns of mortality began to improve for both the most and the least advantaged, but at differential rates. The middle and upper classes seemed to derive the health dividends of industrialization earlier. The mortality rate of the most advantaged improved at a faster rate than the rate of mortality of the least advantaged. The result was that the differences between the most and least advantaged got bigger. However, as time went on, the rate of improvement for the middle and upper classes began to slow, while the rate of improvement for the least advantaged began to increase, resulting in a narrowing of the difference.

This led Antonovsky to suggest that where death rates are relatively high or low, the difference between the most and the least advantaged will tend to be relatively small, but where the rates of mortality are mid range the difference between the most and the least advantaged will be relatively high. Since the publication of these data in the mid 1960s this pattern seems to have evolved still further. For example the gradient in countries like Britain seems to have begun to steepen again over the last forty years or so, and in some countries of the former Soviet block the increase in health inequalities in recent time has been dramatic. One conclusion to be drawn from Antonovsky’s earlier work, combined with the more recent data, is that health inequalities are part of long term social, political and economic trends and are linked to the playing out of policies and historical events and underlying changes in the social structure and the division of labour in society in ways that require an explanation in their own right.
The interesting thing about this is the shape of the curves Antonovsky described. Both extremes are close together and the middle much further apart. One conclusion to draw from this is that it describes a pattern that is linked to some underlying process of modernization/industrialization, and there are some compelling biological (the prevalence of infectious disease, the nature of infant mortality) as well as social (the nature of the housing stock, the appearance of decent sanitation and safe drinking water in particular), set of factors at work. Certainly the chronology of events would lead one in that direction. The other important conclusion is that these data demonstrate namely that inequalities in health are not fixed, but rather are variable at different historical time periods.

One of the more interesting ways of trying to make sense of global type data is to try to evaluate it in the context of data from different spheres. One of the most striking examples of this is in relation to work by Victora and colleagues (2000). They propose the idea of the inverse equity hypothesis. Drawing on data relating to the implementation of child health programmes in Brazil, they note a very similar, almost identical set of curves to that described by Antonovsky, although over very much shorter time horizons. They note that whenever there is a new programme introduced, the children of the better off benefit sooner and to a greater extent than the children of the poorer sections of society. The improvements do impact on the less advantaged but later, and there is an inevitable catching up process. Critically Victora and colleagues argue for the inevitability of this process ceteris paribus. It operates at a much shorter time frame than the kinds of historical epochs which Antonovsky was interested in, but the same pattern emerges. Victora et al., also note that these effects compound one another in the sense that the children of the more well to do are inevitably always in front since the benefits of the next new intervention(s) will have already kicked in, before the poorer cohorts have caught up with last one. So although the overall effect is of health improvement, the constantly repeated cycles tend to reinforce the inequalities giving them the impression of being constant, when in fact they are each the product of successive waves of differential responses to successive interventions.

Several important ideas follow from this. If health inequalities as measured by the differences between different groups are not fixed, but rather constantly reinforced and reproduced, then the search for the distal causes should be in the cycle, or rather in the reasons why inequities reproduce themselves in each cycle in this way. The intermediate determinants are to be located in factors in the delivery of
intervention, rather than in some abstract entity called society, the economy or the political process. These would in turn be mediated by the proximal factors in the life course and the life world.

The key idea here then is the nature of dynamic change, not as interesting historical data but as central to the explanatory process itself.

**Principle 8: Explicating bias**
The last test to be applied relates to ideological or political bias. The problem of values and politics and their particular role in generating knowledge has been a longstanding theme in the social and the political sciences (Weber, 1948). The solution though is relatively straightforward. All knowledge is potentially and actually ideologically or politically biased. The solution is to acknowledge this fact and to seek to make the biases explicit, even if the writer has sought to conceal their own prejudices. This is an imperfect science, but is workable in two stages. First being able to describe any political bias that is inherent in the argument, and then second, to seek to determine whether the political biases have influenced the interpretation of and the selection of the evidence.

It must be acknowledged that values determine the way knowledge is constructed and generated in the first place and in the ways that it is interpreted and used. This explicitly recognises that knowledge and knowing can never be value free. However, the way to move through the problem of values is not to try to produce value neutral science but to aim to make values explicit, to the reader, the author and the user. Recognition must be made that although knowledge and knowing are socially constructed, it is still possible to develop knowledgeable understandings of physical and social phenomena and that the search for better understanding is a worthwhile and important goal of science.

Because knowledge is socially constructed bias will be an inherent part of any knowledge based activity. The overarching goal in reaching understanding of phenomenon is the reduction of and the demonstration of the reduction of bias. In whatever field of knowledge we are working our aim should be to uncover bias and manage it. This is equally true in detecting possible sources of bias in trial data or meta analysis as it will be in interpreting the results of qualitative investigations or the evaluations of policy programmes. That knowledge is a social construct is not denied, but that knowledge is the most valuable tool for tackling health inequalities.
and understanding their social determinants is axiomatic. Knowledge is better than prejudice and is critical to making progress. It is imperative to expose political and ideological biases in the selection of and interpretation of evidence.

3. Into action: Translating evidence for policy and practice

It must be recognised that in public health and related interventions context, culture and human behaviour and social differences in the population play a greater mediating role than in clinical interventions and that therefore different forms of data and evidence will be called into play, external validity will be inherently problematic and the time from intervention to outcome will generally be long term (Briss, 2005). Evidence is an essential but not sufficient basis for policy action. Several other ingredients besides evidence are involved in the policy-making process, including political will, transferability of evidence into appropriate social strategies, and scalability into different contexts and settings. The policy-making process is often poorly understood by researchers (Petticrew et al., 2004; Whitehead et al., 2004; Lomas et al., 2005).

It may be argued that focusing on the evidence base and the individualised interventions which typically populate it has the effect of ignoring the more important macro level determinants of health and the degrees to which inequity is tolerated and sustained through policies at national and local levels. Inequalities in health reflect wider inequalities (and therefore decision-making across a wide policy spectrum). At a micro level, too, reliance on the evidence base works against a recognition or assessment of the effects of synergy and also underestimates context-specific aspects. In order to obviate these problems there are a number of further questions which need to be asked of the evidence (Kelly et al., 2004).

An explicit attempt should be made to assess the transferability of interventions from one setting to another. For example, are the interventions feasible, culturally acceptable and appropriate, and possible to scale-up in developing countries? What are the financing, budgetary and institutional implications of national interventions and of scaling up local interventions? What factors or processes affect change in countries and systems? Under what conditions are policy/programme changes associated with improvements in population-level health and greater equity in health? Such factors and their relative contribution to change should be analyzed.
The MEKN will work towards proposing approaches for using the evidence on integrating the SDH and health equity goals into national and global policy and planning. One particular approach that may be helpful in articulating the processes involved in effective approaches is 'Programme Evaluation', put forward at the expert meeting on measurement held in March 2005. Programme evaluation (PE) in this respect is defined as a systematic set of practices to improve and account for public health actions, and to forecast a range of 'plausible futures' stemming from policies. Evaluation is a driving force for planning effective public health strategies, improving existing programmes and demonstrating the results of resource investments (Milstein & Wetterhall, 1999). The foundation of PE consists of a well-described sequences of steps (engaging stakeholders, describing the programme – including the use of logic models, focusing the evaluation design, gathering credible evidence, justifying the conclusions, and ensuring the use and sharing of lessons learned) that has been set forth by the CDC’s Programme Evaluation Framework (Milstein & Wetterhall, 1999), the Community Toolbox (http://ctb.lsi.ukans.edu/ctb/c30/ProgEval.html) and elsewhere. Beyond formal evaluation systems, PE also encompasses a wider spectrum of methods and approaches (such as system dynamics simulation models) to observe and interpret programmes, and to stimulate further observations.

Within its sequences of steps, PE subsumes virtually all of the methods and tools considered during the meeting and described above. Thus, planning for action within a PE framework incorporates systematic reviews of previous research, Health Impact Assessment, as well as additional steps such as power mapping and eliciting public opinion. The real utility of PE may lie at the country engine level.
References


