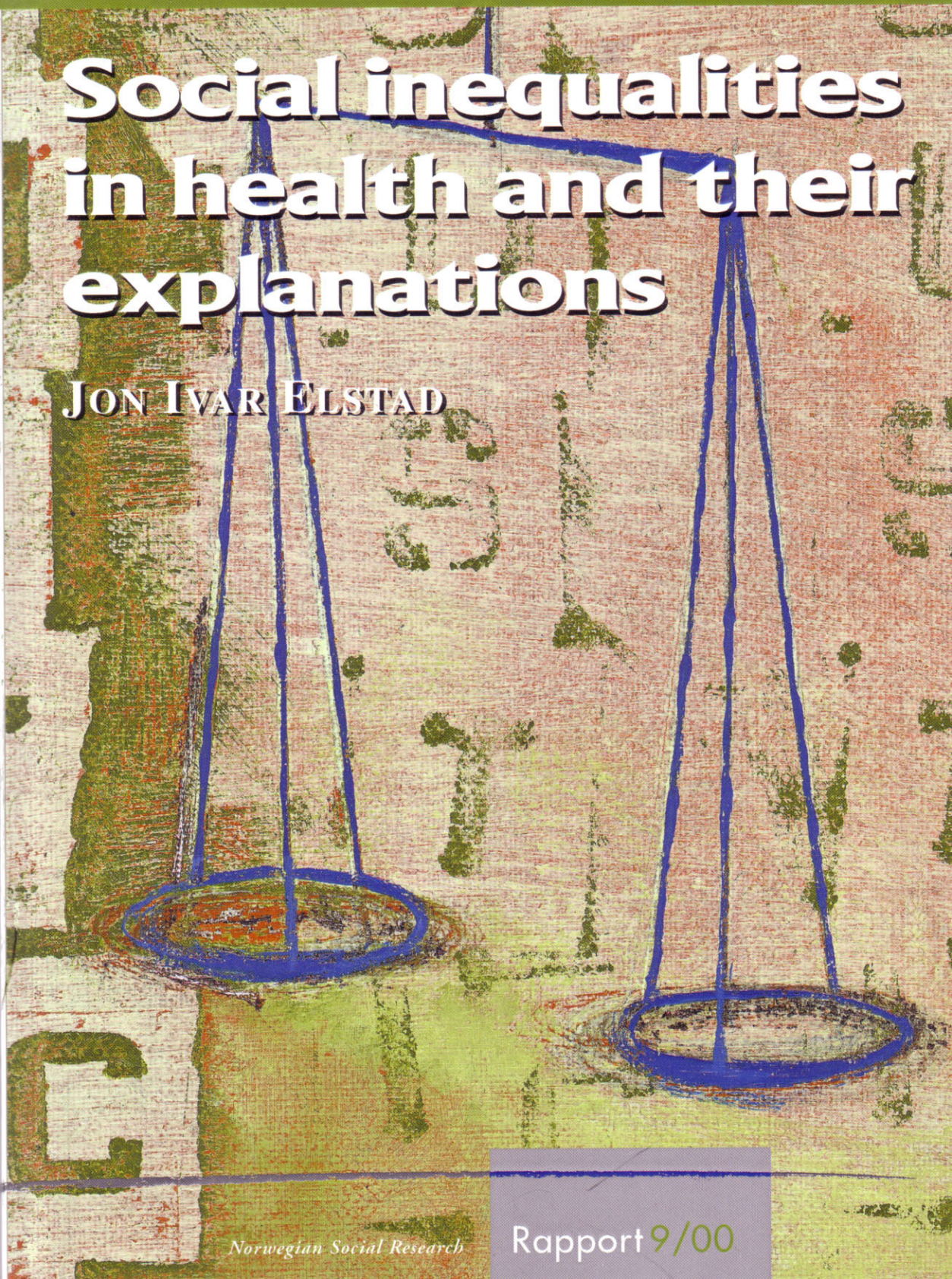
N O V A

Social inequalities in health and their explanations

Jon Ivar Elstad



Norwegian Social Research

Rapport 9/00

Social inequalities in health and their explanations

JON IVAR ELSTAD

Norwegian Social Research – NOVA
NOVA Rapport 9/00

NOVA - Norwegian Social Research is a national research institute under the auspices of the Ministry of Education, Research and Church Affairs, and basic funding is provided by the State.

The main objectives are to undertake research and development projects aimed at contributing towards a greater knowledge of social conditions and social change. The institute shall focus on subjects such as living conditions, quality of life and the life course, as well as on the services provided by the welfare state.

ISBN 82-7894-096-7
ISSN 0808-5013

Desktop:	*Torhild Sager*
Cover illustration:	*ArtVille/AmyDeVoogd*
Print:	GCS

Applications to be made to:
NOVA – Norwegian Social Research
Munthesgt. 29 • Postboks 3223, Elisenberg • 0208 Oslo

Phone:	(+47) 22 54 12 00
Fax:	(+47) 22 54 12 01
E-mail:	nova@isaf.no
Internet:	http://www.isaf.no/nova

Preface

This thesis is submitted for the degree of Dr.Philos. at the University of Oslo. The idea of making this thesis developed through work during the mid-1990s on some papers which addressed various forms of social inequalities in health. The thesis represents a longstanding interest in the subject. Support from my employers, Institute of Applied Social Research (INAS) and its successor Norwegian Social Research (NOVA), and five months' grant from The Research Council of Norway in 1998, made this work possible, which nevertheless had to be done alongside other projects and tasks.

Many colleagues have provided useful and intelligent help and comments to various parts of this work. Professor Dag Album, University of Oslo, has commented on the voluminous introduction. Senior Researcher Espen Dahl, Institute for Applied Social Science, Oslo, has been helpful and supporting during many phases. At various times, I have sought advice about statistical techniques from Professor Torbjørn Moum and Professor Arne Mastekaasa, both at the University of Oslo. The five papers which are included in the thesis have profited considerably from the comments of a number of anonymous referees. The Nordic network on Social Variations in Health has been an inspiring milieu. Others could also have been mentioned.

Translations of quotations from publications in Norwegian or other non-English languages into English have been made by the author.

Oslo – NOVA
May/October 2000

Jon Ivar Elstad

Content

Introduction: the research field of social inequalities in health

That health, illness and mortality are associated with location in the social structure, is old knowledge. The pioneer Norwegian sociologist Eilert Sundt, who investigated mid-19th century social conditions, found mortality differences both between the ecclesiastic divisions of that time and between the social classes – "the property class" and "the working class". In the capital Kristiania (now Oslo), he even noted variations within the working class. Health was better among well-to-do working class families than among the families of labourers. Sundt supposed that two different processes could account for the higher level of ill health in lower strata. Illness could be "an important cause of impoverishment", but also a consequence of the "destitute conditions among the poor" (Christophersen 1979: 117-121, 197, 448).

During the early 20th century, such inequalities in health were well known among population statisticians, public health officers, and some social scientists. How to interpret them was disputed. Many believed that variations in levels of living were the self-evident explanation, but others thought that health formed life trajectories and were the reason why some people climbed while others descended in the social hierarchy. Medical circles at that time were often moderately interested in the relationship between illness and wider social circumstances. The germ theory of disease, gaining momentum after the 1880s, led them to a limited biomedical view on health. There were nevertheless exceptions. The Norwegian physician Karl Evang, Director-General of the Norwegian health services for a long period before and after 2nd World War, remembers how his colleague Olaf Scheel shocked his peers during the 1920s by declaring that "the tubercle bacillus is *not* the cause of tuberculosis". Scheel inferred this because it was found that the majority of the population was carriers of the tubercle bacillus, but many of those infected showed no signs of tuberculosis. Tuberculosis did however vary "with social class and with region", and "the lower the standard of living, the higher the frequency of tuberculosis" (Evang 1974: 100). The lesson was apparently that the germ itself could not explain tuberculosis. The germ was a necessary, but not a sufficient, cause for becoming ill with tuberculosis, and vulnerability to the germ depended on social and material circumstances.

At least since the 1960s, health inequalities have attracted considerable attention among social scientists. Several Norwegian researchers took part in this trend (Blix 1974, Allern 1974, Ringen 1976, Kristofersen 1979, Andersen 1980, Elstad 1981). Today, social inequalities in health has grown into a large research area, pursued by medical sociology, but also by medical psychology, medical anthropology, medical geography, social medicine, social epidemiology, and public health. A major impetus for this boom was the publication of the British *Black Report* in 1980 (DHSS 1980, Townsend and Davidson 1982). The most startling finding in this report was that the socioeconomic mortality differences in England and Wales did not only persist after the introduction of the National Health Service. They seemed to increase: "the lack of improvement and in some cases deterioration of the health experience of the unskilled and semi-skilled manual classes (class V and VI) relative to class I throughout the 1960s and early 1970s is striking" (DHSS 1980: 198-199).

Signs of increasing socioeconomic mortality differentials have also been found in Norway (Dahl and Kjærsgaard 1993, Borgan 1995). The information that health inequalities apparently were becoming more drastic was received with astonishment. The expansion of research activity was doubtlessly inspired by this alarming indication that Western welfare states developed differently from the visions behind them – not towards more social equality, but rather towards more inequality.

During the 1990s a large number of studies investigating health inequalities and their associations with social structure and social processes have appeared. A Dutch research institute located 803 articles on social inequalities in health, published in scientific journals in English and Dutch, during the four-year period 1993-1996 (Whitehead 1998). During the two preceding four-year periods, 760 and 633 articles were found. The main centres of these international endeavours are in the U.K. and the U.S., but also Dutch and Scandinavian contributions are numerous. Health inequalities are often a central theme at international conferences on medical sociology, social medicine, and health policies. Only during 1997 and 1998 three prominent journals within this field published special editions addressing social inequalities in health (*Social Science & Medicine*, Vol. 44, No. 6, 1997, *Sociology of Health & Illness*, Vol. 20, No. 5, 1998, and *Millbank Quarterly*, Vol. 76, No. 3, 1998). At present, several multinational projects are studying these issues. The Scandinavian project "Social variations in health: Nordic comparisons and changes over time" addresses for example whether social change during the 1980s and 1990s has influenced the pattern of health

inequalities (Lahelma 1997). Another example is a newly launched program, "Social Variations in Health Expectancy in Europe" (European Science Foundation 2000), which involves researchers from many countries and which aims at making progress in *explaining* social variations in health.

Thus, the subject of this research field is the systematic patterns of health inequalities found in practically every society where suitable statistics are available. The major research interest is in *socioeconomic* health differentials – the systematic tendency that the higher people are located in the social hierarchy, whether measured by social class, occupational status, educational level, or income, the better is average health. But also other no less distinct patterns of health inequalities are investigated. It is repeatedly observed that married people are of better health than previously married. In most countries, there are persistent patterns of geographical differences in health. Ethnic minorities are generally of less good health than the majority population. Employed people have less health problems than the unemployed and the early retired. And so on. New documentation of such systematic patterns of health inequalities comes every year (for instance Illsley and Svensson 1984, Fox 1989, Jozan 1989, Navarro 1989, Arber and Lahelma 1993a, 1993b, Smaje 1995, Hemström 1996, Kunst 1997, Cavelaars 1998, Nazroo 1998, Elstad 1999a, Lahelma et al. 1999, Dahl and Elstad 1999, Schalick et al. 2000, Rahkonen et al. 2000). The theme of this thesis is these systematic, persistent, and socially patterned health inequalities, and their explanations.

1. The thesis: main topic, overview

1.1. What explains social inequalities in health?

1.1.1. The main topic

The main topic of this thesis is how we can *explain* these social inequalities in health. What are the processes which generate such inequalities? What approaches, explanations, and theories exist for accounting for them? What do empirical studies indicate as regards their relevance? Is there one explanation which appears to have more credibility than others? Are we able to produce a comprehensive theory of these inequalities?

Descriptive studies in Norway, in other parts of the Western world, and wherever we have relevant data, have repeatedly demonstrated typical patterns of health inequalities associated with the major social structures of society – not only the socioeconomic structure, but also the marital structure, the geographical structure, and several other social structures.

How are they generated? What types of processes are involved? The main purpose of this thesis is to contribute to the understanding of this question.

1.1.2. A sociological and an interdisciplinary theme

Both medical and psychological views on health and illness, disease and mortality will be addressed in this thesis. But as Freund remarks: "... each society, in producing its own way of life, produces its own way of death" (Freund 1982:3). Neither medicine nor psychology can, by themselves, explain social inequalities in health. These inequalities are closely associated with society's economic organisation, its distribution of material and immaterial resources, its modes of interpersonal relations, and its social mobility patterns. The production of health and health inequalities is closely related to the features of social life. Thus, the issue belongs to the sociological discipline. But also medicine, psychology, and other disciplines besides sociology are important contributors to these studies, and sociology must interact with these other disciplines when trying to understand how health inequalities emerge.

1.1.3. The reasons for social concern

The existence of a large research field occupied with these health inequalities is in itself a sufficient reason for choosing this topic. The social concern these

inequalities generate is a further reason. Health inequalities in Norway and similar countries influence the life chances and life quality of the population in many ways. These inequalities are a reminder of the *lack* of social equality in present-day societies. They indicate that there are processes which systematically and persistently make certain population categories disadvantaged as regards health. These health inequalities are a problem not only for the parts of the population which suffer from particularly high levels of ill health, but also for population health in general. If health in all population categories was equal to the most advantaged population groups, the average population health would be improved. The values of social equality, as well as the values of population health, are threatened by social inequalities in health.

This thesis is primarily oriented towards the understanding of health inequalities in Norway and similar countries. The problem is much more serious when seen in a world perspective. Data on infant mortality illustrates this. In the 1990s, 150 out of every 1,000 new-born babies died in Mozambique before the age of one, as against five in Norway (Statistics Norway 1998a: table 536). But although the health inequalities in the world system are more alarming, also the health inequalities in the affluent part of the world call for social concern, and they are a public issue in many countries.

1.2. Overview of the thesis

Included in this thesis are five papers published in English-language journals which are important arenas for the international discussion on social inequalities in health. These papers address different questions, but their common denominator is that they all belong to the research field of social inequalities in health, and each of them try to contribute to the clarification of the processes which generate these inequalities.

The other part of the thesis is the more general discussion in this and the following nine chapters. The aims of these chapters are broad. The purpose is to make *a bird's-eye view* of the research field of social inequalities in health. The intent is not primarily to advance a particular view of how health inequalities *should* be explained, but rather to make a comprehensive review of the concepts which are employed, the themes which have attracted interest, and the explanations and theories which are discussed. I want to outline various problematics involved in the explanations of social inequalities in health. Moreover, I want to provide the information required for understanding the relevance of the papers and how they are located in the research field.

Chapter 2 asks: What are social inequalities in health? Often, social inequalities in health are identified with socioeconomic health differentials, but research has also addressed other types of health inequalities. The chapter argues for a delimitation of the research field which comprises all these types of health inequalities. The chapter also discusses the definitions of health used by the research field, which have both sociological and medical origins.

Chapters 3 – 8 describes the explanations of social inequalities in health proposed by the research field. The introductory Chapter 3 argues that explanations are usually formed by combining views on how health is produced with propositions about social processes. Explanations will normally operate with a notion of one or some dominant health determinants which interact with features of society, i.e., with social processes related to society's economic organisation, power distribution, inequality structures, and social relations. Chapter 4-8 examine what I consider to be the five main approaches: the *artefactual* explanation (Chapter 4), the *materialist/structural* explanation (Chapter 5), the *behavioural/lifestyles* explanation (Chapter 6), the *psycho-social* explanation (Chapter 7), and the *social mobility* explanation (Chapter 8). The artefactual explanation addresses primarily questions about the research process, while the four others may be called substantive explanations, as they are primarily oriented towards the *real* processes in society and how they generate health inequalities. Each of these four substantive explanations provide a relatively coherent understanding, different from the others, of how health determinants and social processes interact in society and thereby produce health inequalities. Chapters 4-8 outline the basic features of these explanations, give some remarks of the history of them, describe main empirical themes, important subvariants, and some of the problems and critical questions the explanations have encountered.

In a way, these chapters constitute an introductory text which tries to cover all major attempts which have been put forward in order to understand how social inequalities in health emerge. What is said in these chapters is of course parallel to what is found in many other works. In my view, however, earlier publications have not offered such a relatively short, but nevertheless comprehensive, overview. There exist broad descriptions of the field (see for instance Townsend and Davidson 1982, Evans 1994, Dahl 1994b, Vågerö and Illsley 1995, Marmot et al. 1995, Nettleton 1995, Wilkinson 1996, Bury 1997, Macintyre 1997, Annandale 1998, Robert and House 2000), but I think they are either too restricted in their coverage or too focussed on specific theories. A more all-round overview was therefore judged to be useful.

Moreover, these chapters serve as the background for placing the five papers within the research field. These papers have, as stated earlier, different subjects, but they all address themes which are discussed by health inequality researchers. In Chapter 9, the papers are summarised, and their findings are connected to the main topic about how health inequalities emerge.

This leads up to the final chapter 10. Here, it is underlined that at present, we hardly have a *solution* to the question about how social inequalities in health are generated. This thesis has, hopefully, contributed to the clarification of the problem and brought some relevant empirical evidence. Social inequalities in health, as well as their explanations, develop over time. Social change modifies both the health determinants and the social processes which create these inequalities. The explanations outlined in Chapters 4-8 are the explicit or implicit presupposition for a large number of empirical studies. A general result from these studies seems to be that neither of these explanations can be deemed obsolete, but at the same time neither of them represents, in themselves, a sufficient understanding. At present, the research field attempts to integrate the various explanations in a more comprehensive framework. Two such attempts are briefly discussed – the *life course approach*, and the *community approach.*

2. What are social inequalities in health?

2.1. Definition

When longevity differs markedly between the wealthy and the poor, we readily classify this phenomenon as social inequalities in health. But is women's lower mortality, compared to men's, also included in this concept? If older workers have more health problems than single parents – what is this? Spontaneously we reject the relevance of such a comparison, but why? And if health inequalities arise from foolish behaviour, such as fast driving or overeating, should these inequalities also be included?

In order to delimit the research field of social inequalities in health, I will propose the following definition: *Social inequalities in health are any type of persistent and important differences in aggregated health between social positions in the same social structure(s).* In the following, I will discuss why this definition is advanced. Its aim is to encompass the research activities which I believe should be viewed as part of the same research area. The meaning of the key terms of the definition – social positions, any type of aggregated health differences, and health – will be discussed. An important part of the definition is health, and the latter part of the chapter will address the concepts and indicators of health used in research on health inequalities.

Initially, a few points can be treated briefly. Social inequalities do not refer to differences between individuals, but to differences in *aggregated health* between population categories. Within any large population category one will always find substantial individual health variation. Social inequalities emerge when these individual differences, added together, manifest systematic differences between population categories. "Social" is in this sense the opposite of "individual". When we speak of social inequalities we are not comparing individuals, but collectivities of individuals.

Moreover, the health differences should be *persistent* and *important.* "Persistent" means relatively enduring – not unchanging, but stable in the sense that the inequality patterns exist for a fairly long period, say, at least ten years. "Important" requires that the inequalities imply differences in life quality and life chances which have social significance. What types of differences that qualify according to this criterion is a matter of debate. If average longevity differs by a few months, we would judge the difference – no matter if it is significant in statistical terms – as trivial. When life expectancy is 7.3 years

shorter for men residing in the central eastern parts of the Norwegian capital Oslo than for men residing in the richer western suburbs (Rognerud and Stensvold 1999:82), these differences are important. In many instances, it is of course more difficult to decide whether differences are important or not.

2.2. Social positions

2.2.1. Not only socioeconomic inequalities

Often, social inequalities in health are identified with inequalities pertaining to social class and the socioeconomic stratification of society. I will however propose that all types of systematic health inequality patterns, related to every major social structure, should be included in the research area. Therefore, the definition does not point only to socioeconomic differentials, but to health inequalities between "social positions within the same social structure(s)".

That the system of social stratification is a major determinant of people's life chances has been a dominant theme in sociology, from Marx, Durkheim and Weber to Giddens (1973) and Bourdieu (1984). It is therefore not surprising that health inequalities have most often been seen in terms of this division. To what extent social stratification continues in present-day Western societies to have an all-pervasive influence on people's lives, has however become a disputed issue (cf. for instance Esping-Andersen 1993, Pakulski and Waters 1996, Crompton 1996, Higgs and Scambler 1998). This suggests that research on health inequalities should have a broader perspective and not be oriented solely towards socioeconomic differentials.

Moreover, there are systematic health inequalities which cannot be conceptualised in terms of social class or social stratification. Differences according to marital status, residential area, or ethnicity, are not necessarily independent of class, but cannot be identified solely with class inequalities. If social inequalities in health are restricted to socioeconomic differentials, the multiple ways by which health is patterned in society may be overlooked.

Finally, an important argument for including not only socioeconomic differences is that the study of how health inequalities are generated will regularly be conducted by means of the same methods and discussed in terms of the same theories, whatever social structure we focus on. One will often employ the same conceptions of the determinants of health, and discuss the same types of social processes, no matter whether marital status, social class, or geographical health differences are examined. The unity of concepts in all health inequality studies makes it artificial to define the research field only as the study of socioeconomic differentials.

2.2.2. Social structures and social positions

Given that studies of socioeconomic health differentials, health differences between marital status categories, health inequalities according to geographic location, etc., have so many similarities that they belong to the same research field, the question arises: What are the underlying principles for constructing the population categories between which health is compared?

I will contend that the common feature is that they construct population categories as social positions belonging to the same social structure(s).

"Social structure" is a term with different connotations. According to Warriner (1981: 179), "Of all the problematic terms in the sociological lexicon, 'social structure' is perhaps the most troublesome". There are many differences between the conceptions of social structure given by, for instance, Lévi-Strauss (Leach 1981), Blau (1981) and Giddens (1984).

I will suggest that when the term is used in research on health inequalities, social structure refers to an important *analytical aspect* of social life and the main *social positions* connected with this aspect. An analytical aspect of social life is a particular social relation or social characteristic. When studies address health inequalities related to occupational class, for instance, occupational class refers to the aspect of social life which is related to the person's economic activity, working conditions, and location in the status hierarchies of firms or other work organisations. Social positions within this aspect are, for instance, farmers, workers, white collar, and self-employed, but these positions may be defined in different ways, depending on whether we focus on type of work, authority, qualifications, or property. Taken together, these positions constitute the *occupational structure* of society.

Similarly, when research addresses inequalities related to marital status, one focusses on how people are, or are not, living in intimate relations to another person, and the different positions related to this aspect constitutes the marital structure. When research addresses health inequalities according to educational levels, one focusses on the social structure which is constituted by categories of education. And so on. The common feature is that terms such as occupational class, marital status, labour market position, income level, educational level, residing area, and social background – categories often employed in studies of health inequalities – all point to a particular aspect of social life and construct a social structure in terms of the main social positions within this aspect.

Social structure and social positions are linked to each other. The structure is the totality of the social positions within the structure, and the bond between social positions is that they refer to the *same* structure. – This is why comparisons between older workers and lone parents, or between children

and people from Northern Norway, will be seen as irrelevant for research on health inequalities, as they obviously refer to different aspects of social life: workers and lone parents are not social positions in the same structure.

2.2.3. Supplementary remarks

Social structure and social position are analytical constructs. By using these terms, we do not refer to societies or persons as wholes. The social structure we focus on, emerges by abstracting one particular aspect of social life from the totality. Likewise, the social positions arise by "slicing out" certain parts of the persons' characteristics or relations, and by classifying the individuals in terms of these characteristics or relations. Thus, the social position is not a person or an individual, but the collectivity of all persons having the same characteristics or relations as regards the aspect we focus on. The social position will always consist of persons who differ in many ways, but who are grouped together because they are similar as to the specific aspect of social life we want to examine. When we describe the health of a social position, we are not referring to individuals' health, but to aggregated health – usually described in terms of mean values or distribution of health profiles – of all the persons who are classified together in the same social position.

Using this analytical approach, it is easily seen that society has quite a few social structures. But is research on social inequalities in health entitled to construct any type of social structure, by focussing on a particular, even peculiar, aspect of social life, and claim that the health inequalities as regards this aspect are part of the research area? Of course, the research field concentrates on social structures which are meaningful in some sense. We may think analogously to Weber's definition of class, which he constructs in terms of "a number of people (who) have in common a specific causal component of their life chances" (Gerth and Mills 1970: 181). Meaningful social structures are those structures which, in some sense, form an important element of people's lives, life chances, or ways of living. Doubtlessly, a large number of social structures are relevant according to this criterion: structures involving age, gender, ethnicity, social class, occupation, labour market, education, income, wealth, property, family situation, marital status, social background, residing area, geographical location, welfare benefits and pensions, etc.

Research on health inequalities does not necessarily involve only one social structure at the time. Also combinations will be studied – labour market status *and* marital status *and* gender, for instance (are married working women less healthy than married non-employed women?). In such cases, the social

positions which are examined are combined social positions, referring at the same time to several social structures.

A social structure is no constant. When research refers to the "same" social structure when comparing, for instance, health inequalities between occupational positions at two time points, this "sameness" is relative. Social structures are regularly changing in four typical ways: (1) The relative proportions of the different social positions change. As regards the marital structure, for instance, divorced people make up a larger part of this social structure in the 1990s than in the 1960s. (2) New social positions develop within the same social structure. The transformation from industrial to post-industrial societies was accompanied by new social positions in the occupational structure. (3) The circumstances pertaining to specific social positions will always be subject to change. Manual workers are confronted with different working conditions today, compared to some decades ago. (4) Finally, the actual individuals making up the social position will change. Some move to other social positions, new members are recruited. When the "same" social structure is analysed at different time points, we are seldom analysing the same individuals.

2.2.4. Age, gender, ethnicity

The definition implies that studies of age and gender differences are also part of the research field. It may be argued that age and sex categories are biological categories which do not refer to social structures. Doubtlessly, biology is not irrelevant for the health differences in the age and gender structures. The health deterioration which is part of the ageing process is hardly only a social product. Health conditions *must* to some extent vary between men and women – if not for other reasons, at least because different body organs imply that some disorders can only affect females, others only men.

Nevertheless, age and gender categories are social positions. Age categories – adolescence, middle age, the elderly – refer to people in particular social situations who are influenced by typical expectations and norms and are part of typical material and social contexts. These circumstances are *social*, and the age structure is in this way also a social structure.

That social change modifies how the ageing process affects health supports this view. The life expectancy for the typical fifty-year old individual has changed markedly during the last century. Similarly, the health differences between men and women depend on social circumstances (Elstad 1999a). Medical studies often suggest that gender health differences may also reflect "pure" biology. One cannot deny the possibility that if society was stripped of all gendered social structures, some sex differences in health would anyway

survive. However, as there are compelling evidence that social circumstances affect gender health differentials, gender must be considered primarily as a social structure in health inequalities research.

Ethnic positions are social positions in the same manner. In medical epidemiology it is often suggested that health differences between, say the Lapps and other Norwegians, or between blacks and Caucasians in the U.S., are reflecting biology, i.e., different genetic distributions. It is of course possible that a population category which is denied intermarriage with other population categories will gradually develop distinct genetic profiles with some health consequences. It has been argued that the higher prevalence of diabetes among Indians has a genetic component (McKeigue 1997: 94). However, to my knowledge very few genetic differences between broader ethnic categories with *important* consequences for population health have been demonstrated.

Nevertheless, in one sense the study of health inequalities between categories of gender and ethnicity is different. While people may change social positions as regards most other social structures – the income structure, the occupational structure, the family structure, etc. – sex and usually ethnicity are fixed aspects of people's situation. Mobility between social positions may be involved when health inequalities in other social structures emerge, but mobility is usually not relevant when gender and ethnic health differences are examined.

Although age, gender, and ethnicity are social structures, they will nevertheless often be treated in a special way in health inequality research. It is common to examine the two gender separately. The reasoning behind this is often unclear, as the determinants of health are often supposed to be the same for both sexes. However, as systematic gender differences in health (higher male mortality, higher female morbidity) are repeatedly found, one will often study men's and women's health apart. Furthermore, age is regularly employed as a control variable. Since the categories we compare often have different age compositions, and age is markedly related to many (but not all) health indicators, studies commonly adjust for age in order to demonstrate the "true" differences in the social structure which is analysed.

2.2.5. Social structures and social theory

Thus, the research field is characterised by its interest in health inequalities as regards a number of aspects of social life – the socioeconomic aspects and many others, among them also age, gender, and ethnicity. The common feature is that the research field addresses health inequalities between population categories which are social positions in the same social structure(s).

An implication of this is that the research field is not defined in terms of a specific theoretical approach, i.e., one specific way of constructing population categories. Researchers within the field employ a number of ways of defining their population categories. Different versions of the social class structure, the residential structure, or the family structure, rooted in different social theories, are found within the research area, as well as pragmatic classifications provided by official statistics without any explicit theory behind. The definition encompasses all these varied ways of constructing social positions, and its intent is to be *neutral* as regards social theories.

However, is it really neutral? It views social structures as a number of co-existing aspects of social life. It might be said that this diverts attention away from the social whole, from societies at large, and also from individuals as whole beings. The question about social inequalities in health is formulated as a question about health inequalities pertaining to a specific aspect. If we advocate a general theory of society, imagining it as a totality founded on some basic elements (say, the relations of production, or the relations of gender), and claim that all kind of divisions in society are interconnected and have developed from these basic determinants, we might feel that the analytical approach is biased: it disconnects the social structures from the social whole.

However, when studying health inequalities empirically, one will, I believe, eventually be pressured to translate grand theories into specific propositions about how the majestetic social forces influence the health distributions in specific social structures. At this point, at least, I believe that the approach proposed here for delineating population categories will be applied, whatever theoretical foundations one starts with.

2.3. Any kind of health inequalities

The definition states that *any* kind of systematic and important health differences between social positions are examples of social inequalities in health. To clarify why this point is included in the definition, I will discuss some viewpoints presented by Margaret Whitehead (1992).

Whitehead makes a distinction between *inequality* and *inequity*. These terms are frequently employed as synonyms, but Whitehead emphasises that the first term should denote differences "in a purely mathematical sense", while the second term implies, in addition to differences, also a moral and ethical dimension. Health inequities are those differences which are "unnecessary and avoidable, but in addition are considered unfair and unjust" (1992: 431). Accordingly, health inequities are not only *factual* differences between

population categories, but differences which need not exist, which could be avoided, and which public opinion would condemn. To decide whether inequalities are also inequities requires therefore both a valid understanding of how the inequalities have been generated, and a moral or political judgement as to the fairness of the processes which produce the inequalities.

As a guideline for political action, the distinction between pure inequalities and inequities is reasonable. In passing, it can be noted that Whitehead's paper appeared in the context of discussions within the World Health Organisation, a political body which, of necessity, should define its task as fighting unnecessary and avoidable health inequalities.

However, the research field of social inequalities in health is not restricted to inequities. It consists of studies addressing different types of health inequalities, and both "fair" and "unfair" processes are proposed as the explanation of these inequalities. Although inequities have high priority, the field is also addressing inequalities which could eventually be deemed as necessary, unavoidable, and fair. Whether pure inequalities or inequities are addressed, the studies are part of the same research field, as their theme is health inequalities between social positions and their methods and concepts are, by and large, the same.

Generally, the social positions which are more severely afflicted by health problems, are also less privileged as regards other types of resources and assets. However, there are exceptions. The elderly are more afflicted by ill health, but as regards material wealth their situation is probably better than the situation of younger people (at least in Norway). Breast cancer seems to be more frequent among women who are higher up on the social scale (Link et al. 1998, Vatten 1999). There is hardly any reason that such "deviating" patterns of health inequalities, which perhaps are not inequities in Whitehead's sense, should not be part of the research area.

Another reason for not defining the field in terms of inequities, is that the processes behind are often insufficiently clarified. It is not easy to classify them as avoidable or unavoidable, or fair or unfair. As to the typical health inequalities in affluent countries, the *disputed* issue is exactly what types of processes can account, to what degree, for the observed health inequalities.

The criteria unnecessary, avoidable, and unfair raise questions which, to a large extent, cannot be resolved by research itself. To avoid a certain type of health inequalities is not only a question about what kind of processes that generate these inequalities, but also a question about how large resources are we to invest in avoiding them, and what types of social change we imagine as realistic. As to fairness: opinions are divided. Whitehead judges downward

social mobility in consequence of health problems as unfair, but not everyone will agree. Research can investigate how health influences mobility, and to what extent this contributes to the pattern of health inequalities. To draw moral and political conclusions will, in the end, be the task of public debate (where researchers of course have a special responsibility for voicing their opinion).

2.4. Health

The focus of this section is concepts of health, not how health is generated. What health *is* and how health is *produced* cannot always be separated, but here I will concentrate on the former of these topics. The aim is to survey how health is understood in research on health inequalities, and what types of health indicators the research field employs.

Medicine is often considered the expert discipline as regards health. A medical view on health is frequently applied in health inequalities research. However, there are also sociological viewpoints on health, regularly used in this research. I will outline the differences between these two conceptualisations, but also discuss how they often coincide in practical research.

2.4.1. Health – sociological views

In the latter half of the 20th century, main theoretical trends within sociology, such as functionalism, phenomenology, symbolic interaction, marxism and conflict theory, developed particular approaches to health sociology (Gerhardt 1989). The differences between these trends will not be treated here. The purpose is to sketch the common concept of health which, by and large, is present in all these trends.

One starting point for the sociological views on health is sociology's interest in humans as *social actors*. For Weber, for instance, social action was a central topic. Social actions are meaningful behaviours oriented towards other actors (Ritzer 1975: 91). Social action refers to the individual's consciousness of being located within a social environment, to his/her purposive behaviours and actions directed to other persons, and to the individual's integration in and interaction with his/her social milieus.

However, social action presupposes that the body and the mind are able to perform, to function. The health preconditions for social action were not made explicit by Weber, but, as Gerhardt points out: "physical or mental health is necessary if actors are to take part in subjective reciprocal meaning formation" (Gerhardt 1989:xi). The *capacity to act* depends on the state of the body and the mind. One sociological approach to health is therefore to define it as the bodily

and mental preconditions for participation in social interaction and in society. Correspondingly, failing health refers to bodily and mental states which incapacitate people's possibility to act, meaningfully, towards others.

The health preconditions for social action are subjective and objective. In subjective terms, health is the experience of the state of one's body and mind, the feeling whether this state is satisfying or not, and the perception of control over the body and the mental apparatus. Objectively, health is bodily and mental resources for social action. Somatic disease may impair the body, while physical strength and endurance increase the capacity to act. Likewise, psychological problems constrain the mental resources for social participation, while mental balance and psychic energy enhance the possibilities for social interaction and for making oneself influential in the social world.

In this view, health is conceptualised as a resource for the individual, and health is seen in *instrumental* terms, as an asset which enables the attainment of other values. Somewhat different is the sociological approach which emphasises the inherent value of health. The World Health Organisation's definition of health – that health implies not merely the absence of disease and infirmity, but is "a state of complete physical, mental and social well-being" (Gerhardt 1989: 304, Evang 1974: 34) – suggests the value of health primarily as a *welfare* component. Instead of stressing humans' capacity to act, the WHO definition points to enjoyment of well-being, and emphasises passive, rather than active, aspects of health.

The WHO definition is not dissimilar from how health is conceptualised by theorists reflecting on the "postmodern condition". Bauman argues that while the body in the heydays of industrialism first of all was related to work and soldiering, i.e., as instruments for production and war, "the postmodern body is ... a receiver of sensations; it ... digests experiences; capacity of being stimulated renders it an instrument of pleasure" (Bauman 1998:226). Fitness as well-being, rather than a precondition for making an impact on the outer world, comes in focus in postmodernist ideas of health.

Health as a resource for action and health as well-being constitute two somewhat different sociological conceptions of health. They are however not mutually excluding, and the health indicators we use will often coincide whether resources or well-being are seen as the essence of health.

2.4.2. Contrasts between medical and sociological views

The sociological interest in health is rather new. There are forerunners, but medical sociology has mainly developed during the last fifty years (Gerhardt

1989, Bury 1997). Historically, medicine is the major discipline oriented towards the study of health, illness and disease. Various conceptions of health have been present in the western tradition of medicine, which dates back some 2,500 years. The main medical view of health as of today is primarily a product of the growth of scientific medicine during the last centuries, and it has certain characteristics which differ from the sociological conceptions.

The roots of the medical view can, in a stylised way, be found in the contrast between two Greek gods of antiquity – Hygeia and Asclepius (Dubos 1995, Renaud 1994, Evans and Stoddart 1994:29). Hygeia was the god of wise living, prescribing how saneness and a long life could be secured. Asclepius was the *healing* god, famous for his knowledge of how to defeat disease. Although Hygeia never was out of sight, Asclepius became the foremost ideal for medicine. This follows from the role of the medical profession. The tasks of physicians were, and are, primarily to cure sick individuals (Rose 1985: 32, Berg 1987: 14). Rather than being preoccupied with health, doctors are oriented towards illness and disease, and how deviant functioning of the body and mind can be restored.

Thus, a source of sociology's and medicine's different views on health are the two professions' different social practices (cf. Habermas 1969). The interest of sociologists was to understand health in its connections to people's experiences, social relations and social actions. The interest of physicians was to understand *failing* health, i.e., disease, and, above all, how it could be cured.

This curative orientation has been fundamental for the development of medicine. Physicians took part in the scientific developments which gained momentum during the Renaissance (Samson 1999). Searching for the essence of disease, they looked into the body, using more and more sophisticated instruments and gaining more and more insights into human anatomy, physiology, cells and bodily processes at the molecular level. The medical paradigm became *biomedical* and *reductionistic*, viewing illness and disease not in terms of the experiences of the sick, but in terms of the details of the hidden processes inside the body (Berg 1987, Nettleton 1995: 3, Annandale 1998: 6). Diagnoses became a central category. As medicine progressed, the ways of falling ill were evermore precisely classified. The first international classification (1855) provided 138 diagnoses. The present International Classification of Diseases used by Statistics Norway (ICD-9) names close to 10,000 disease entities, and an even more detailed classification is soon to be implemented (Statistics Norway 1993: 7, 3).

Given this historical background, some contrasts between sociological and medical views of health can be outlined. The following points to tendencies

which emerge from the different social practices of sociologists and physicians. In practice, however, many sociologists have been strongly influenced by medical conceptions. A major trend in medical sociology has been termed sociology *in* medicine, defined as "the application of sociological concepts, knowledge, and techniques in efforts to clarify medical and social-psychological problems in which medical workers are interested" (Kendall and Reader 1972: 2). On the other hand, many physicians have been inspired by psychology and sociology. It has been contended that the biomedical stage in medicine was particularly dominant from about 1850 to 1950, but that medicine later became more open to other influences, not least from the social sciences (Svensson 1993: 15). Nevertheless, biomedicine is dominant even today, and therefore it is relevant to characterise the contrasts between sociological and biomedical views.

(1) Biomedicine considers individuals outside of their social milieus. As the standard is the "normal" human organism, devoid of bodily abnormalities, medicine can classify a person as sick even if the person generally performs well in his/her social life. The sociological view will, on the other hand, regard the individual as localised in a social context. Health cannot be isolated from the social surroundings, but must be seen in relationship to the performance society expects from its members.

(2) Biomedicine focusses on specific diseases, i.e., specific "faults" pertaining to the body and the mind. Sociology leads up to a holistic view, regarding humans as whole persons, and health as the unity of bodily and mental states.

(3) Biomedicine is oriented towards deviations from normality and defines health in terms of the absence of specific symptoms, illnesses and diseases. Biomedicine is primarily occupied with the negative aspects, with failing health, and with the discontinuity between the healthy and the sick. The sociological view implies that health is a continuum, spanning from the very bad to the very good. Positive health, indicated by more resources for action and more well-being, is meaningful within the sociological view, but usually absent from biomedicine.

(4) Biomedicine depends on professional judgement and medical expertise, in order to decide whether the person is sick or not. The sociological view points to the individual's performance within his/her social milieu, to lay experiences of health, and to judgements by the person him/herself and by persons surrounding the individual. Instead of medical judgements, sociology emphasises social definitions of health and illness.

Summing up, sociology points to the *social significance* of health, and judges a person's health in terms of his/her bodily and mental well-being and

resources for social action. The orientation of biomedicine is towards diseases and their cures. In some ways, medicine has a more exact and "scientific" approach to health, however a restricted view with a focus on negative aspects and deviations from normality. The more holistic view of sociology will on the other hand be confronted with definitional problems. Health is conceptualised as mental and bodily resources and well-being. That disease reduces these aspects is clear, and as to failing health, sociological and medical views will often coincide. But what about positive health? Mental and bodily resources, enhancing social participation and well-being, are not only dependent on the absence of disease, but depend also on physical strength, endurance, vigour, vitality, energy, boldness, propensity to act, and even on weight, height, appearance, beauty, intelligence and joyfulness. Are all such mental and bodily characteristics part of the concept of positive health? If not, where should the dividing line between the health and non-health aspects of mental and bodily states be drawn? Sociology has not a clear answer to such questions.

2.4.3. Health indicators employed by the research field

In research on health inequalities, health will be measured by *health indicators*. Both sociological and medical views on health may constitute the basis for these indicators, but not seldom the indicators themselves will be the same, no matter whether a sociological or a medical view lies behind. In the following, the main ways of measuring health will be outlined, and I will suggest how they are related to the sociological and medical views.

Mortality is the most frequently used health indicator, partly from pragmatic reasons: death is a "clear-cut" event (Bury 1997: 55), well registered in countries with developed population statistics. Death represents both sociological and medical views – it is the end of social life, as well as the ultimate bodily pathology. *Cause-specific* mortality is more close to the medical interest in disease entities, but will also have sociological interest. Deaths from murders, accidents, suicides, aids, and many other contagious diseases, are clearly related to special types of social action and interaction.

Historically, mortality has been the supreme indicator of population health. However, in countries such as Norway, deaths have become relatively infrequent among young, and even among middle-aged, people. Eighty-five per cent of all deaths in Norway in 1995 occurred among people aged 65 years and more (Statistics Norway 1998a: table 86). Therefore, it can be argued that the health situation of *the living* is equally, perhaps more, relevant for population health than are mortality statistics (Blaxter 1989: 199).

Morbidity is, within the sociological view, usually represented by illness, often in terms of somatic and mental symptoms, complaints, and pains. Such experiences will influence well-being and the capacity to act. The medical view of morbidity will focus on the diagnoses assumed to underlie illness. Diagnoses and illnesses are associated with each other. However, there is no one-to-one association, and people with the same diagnosis can have widely different experiences of illness and symptoms (Elstad 1998a: 35). As there nevertheless is a statistical correlation between illness as experienced by the person, and diagnoses as revealed by medical examinations, diagnoses can function as an indicator of illness experience. *Self-reported* diagnoses are a useful, although crude, approach to medically defined disease. When people are asked about disease in countries with extensive health services, they commonly report the diagnoses they have been told by their doctors (Bjerkedal and Bakketeig 1975, Blaxter 1989: 207).

Disabilities – the ability to walk, talk, listen, to fulfill job requirements, to receive education, etc. – refer to the possibilities to perform social roles and to participate in social life. It is therefore a direct approach to the sociological view of health. The biomedical interest in disabilities will often concentrate on the underlying diagnoses. Somewhat related to disabilities is *health expectancy* – expected years free from disability and major illnesses. It aims at being a "synthetic overall measure of health" (Sihvonen et al. 1998) and is clearly connected to the sociological view.

Health potential is, according to Wilkinson (1986a: 6), an idea that "still needs to be clarified". It may be conceptualised as bodily or mental traits which are not, in themselves, characterised as illnesses or medical disorders, but which influence the likelihood of later disease and mortality. Understood in this way, health potential is a *latent* characteristic, and therefore difficult to measure. Birth weight, height, previous diseases and injuries, and also childhood socioeconomic circumstances and health behaviours are used as indicators of health potential (e.g. Lundberg 1991, Barker 1992, Power et al. 1996b, Dahl 1996). Such indicators are found to correlate statistically with later morbidity and mortality, but it is not always clear whether this correspondence signifies a causal connection or a spurious one. Variations in genetic equipment may also be supposed to generate variations in health potential. The indicators of health potential are more close to medical than to sociological views of health.

Self-perceived health and subjective well-being are important ways of eliciting people's feelings of a good or bad health status, and their experiences

of general mental health. As the focus is the person's own subjective opinions, such indicators are primarily associated with sociological views on health. They enable measurements of *positive health*, if they are designed with that purpose in mind. Such indicators play also an increasing role in medical evaluations of therapies and hospital treatments. Self-perceived health status is a simple, but nevertheless powerful indicator. Several studies indicate that self-rated general health predicts later disease and mortality more accurately than medical evaluations (Idler and Benyamini 1997).

2.4.4. Health: unidimensional or multidimensional?

The dissimilarities between sociological and medical views of health and the existence of different health indicators raise the question whether health is a *unidimensional* or *multidimensional* phenomenon. The multitude of health indicators reflects that health is a complex phenomenon – not only for researchers, but also for "ordinary" people. Studies of lay conceptualisations of health reveal hardly less differentiated opinions among the population as to what health "means" than can be found among researchers (Fitzpatrick 1984, Blaxter 1990, Bury 1997: 29pp). The variety of lay opinions reflects that in everyday life, "health" is a diffuse notion involved in a plurality of experiences encountered by everyone.

One may approach the question about the dimensionality of health from the *empirical* angle. Using individuals as the unit of analysis, we will often find a statistically significant, but nevertheless moderate, correlation between different indicators of health. Data from Statistics Norway's *Health Survey 1995* give some examples. For the sample aged 24-67 years, correlations (Pearsons r) are + .53 between the (ordinal) scale of self-perceived health and number of diagnoses, + .33 between self-perceived health and an index of mental problems, and + .39 between the mental health index and number of diagnoses. Such correlations suggest that different aspects of health are statistically associated at the individual level, but the strength of the associations is hardly large enough to maintain that the different indicators are only different ways of measuring the same, unidimensional, underlying phenomenon. This corresponds to everyday knowledge that individuals may experience mental distress but are otherwise physically fit, some claim their health is pretty good although their disabilities are manifest, etc.

However, in research on *social inequalities* in health, the main interest is how the relationship between different health indicators appears at the collective level, i.e., when the units of analysis are social positions and not individuals.

More or less the same pattern of health differences between socio-economic positions are usually found for most chronic diseases, disabilities, mental symptoms, and self-perceived health. Marmot et al., for instance, analysed three different data materials from the U.S. and the U.K. and found "similar social gradients for adult men and women in physical and mental morbidity and in psychological well-being" (1997: 901). This does not always imply that different indicators behave in the same way. The direction of the inequalities is however the same. This is also exemplified in one of the papers included in this thesis, which shows relatively similar gradients between occupational classes for such diverse indicators as bodily height, self-perceived health, and number of somatic symptoms, mental symptoms, and medical diagnoses (Elstad 2001).

Furthermore, the pattern of morbidity differences corresponds to the pattern of mortality between socioeconomic positions. Moreover, most specific causes of mortality follow the same inequality pattern as overall mortality. In England, it has for instance been shown that the familiar pattern of mortality differentials between occupational classes appears for no less than 65 of the 78 most common causes of death for men, and for 62 of the 82 most common causes of death for women (Wilkinson 1996: 71).

Thus, when the units are not individuals, but social positions indicating social class, occupational status, and other classifications of socioeconomic location, one usually finds very similar patterns for different health indicators. This suggests that health could be considered as unidimensional at the aggregate level. The same tendential unidimensionality of health appears in other social structures as well, for instance the ethnic structure, the structure of geographical locations, and the marital structure. As to the latter, the health advantages of married, compared to previously married (with never married somewhere in between), are found both for mortality (Mønnesland et al. 1982, Hemström 1996), longstanding illness (Elstad 1996a), and subjective well-being (Mastekaasa 1993a). Accordingly, the disadvantaged social positions usually display higher levels of ill health, and the advantaged positions higher levels of good health, whatever health indicator which is examined.

The reason for this may be that there are some common processes which cause quite varied types of ill health to emerge more frequently among disadvantaged social positions. It has been suggested that this is a consequence of social positions' varying *general susceptibility* to health problems. I will discuss further this concept in the next chapter.

The tendential unidimensionality of health at the aggregate level has some practical consequences for research. It suggests that, when the units of analysis are social positions, chances are high that whatever health indicator we focus on, it will display patterns similar to most other health indicators. Thus, mortality differences will usually also indicate how morbidity is distributed, and the distribution of somatic morbidity will also indicate how mental problems are distributed. However, this rule-of-thumb is no "law", and one must always be aware of the possibility that some specific health indicators may behave rather differently than the general pattern.

Important *exceptions* to this tendential unidimensionality appear when one examines health differences between age categories and between the two gender. The association between health and age is different for different health indicators. Mortality and chronic disease increase as people grow older, but as regards subjective well-being and mental symptoms, the young are generally not less afflicted than the middle-aged. Even more deviating are gender differences. A well-known paradox is that male mortality always is higher, while female morbidity, in particular somatic symptoms and mental distress, regularly surpasses men's (Elstad 1999a). "Women feel sicker, but men die quicker." Thus, the tendential unidimensionality of health, when differences in health are analysed at the aggregate level, does *not* apply as regards the social positions in the age and gender structures.

3. Explanations of social inequalities in health: overview

3.1. The dividing lines between explanations

The purpose of Chapters 3 – 8 is to survey the explanations of social inequalities in health. The aim is to discuss how researchers have tried to interpret the systematic patterns of health inequalities, to point out the main dividing lines between these attempts, to classify the main explanations, and to describe their main themes, problematics and problems.

Although we may find earlier antecedents, it is reasonable to say that research on health inequalities originated primarily in the 19th century. It has expanded in quantity and diversity, especially during the last decades. A number of explanatory alternatives have been produced. They differ according to many criteria: their level of generality, the factors and conditions they focus on, the theories they explicitly connect their explanations to, and their implicit presuppositions. It is not easy to classify all these attempts into neat and mutually exclusive boxes. Studies and researchers will often draw on several theories and mix elements from many sources. Because of this, the dividing lines tend to be blurred. The distinctions between the different alternatives are not always clear-cut, and often hybrid explanations are put forward.

3.1.1. Two aspects of explanations

A suitable starting point for classifying the explanations is, in my opinion, to point out that they usually address *two* questions. Explanations of health inequalities will regularly have two aspects: one which addresses health, and another which addresses society. When explanations refer, for instance, to environments or lifestyles or human biology as ways of accounting for health inequalities, two propositions are really put forward: first, that environments or lifestyles or human biology are very important for health, and second, that society operates in a way which produces significantly different environments (or lifestyles or human biology) between different social positions.

Accordingly, most explanations refer simultaneously to propositions about health and to propositions about society. They are constructed on the basis of notions of *health determinants* influencing the health in the population (or in particular categories in the population), but also on notions

about *social processes* which somehow influence or interact with the health determinants and thereby produce the pattern of health inequalities.

This distinction suggests two different ways of classifying explanations: according to what kind of social processes they operate with, and according to their notions of the main health determinants. This distinction is some of the reason why authors classify what goes on in the research field in different ways. It can be seen when we for instance compare how the Canadian *Lalonde Report* (published in 1974) and the *Black Report* divided the field. The Lalonde Report saw population health as determined by life style, environment, human biology, and the organisation of health services (Evans and Stoddart 1994: 41), implicating that these factors were also responsible for social variations in health. The *Black Report* (DHSS 1980, Townsend and Davidson 1982) saw four main alternatives: Artefact explanations, natural/social selection, material/ structural explanations, and cultural and behavioural explanations.

The artefact explanation is special because its subject is primarily the *research process* and its implications for how the picture of health inequalities emerges. Apart from this, the different dividing lines the two reports propose can be interpreted as reflecting that the Lalonde Report is primarily oriented towards classifying health-determining conditions, while the Black Report is more oriented towards classifying social processes.

A similar comment can be made when we compare how Whitehead (1992: 432) and Hertzman et al. (1994: 76-78) make their classifications. Whitehead proposes that social variations in health may be due to seven processes: "(1) Natural, biological variation. (2) Health-damaging behavior, if freely chosen.... (3) The transient health advantage of one group over another when that group is first to adopt a health-promoting behavior (as long as other groups have the means to catch up fairly soon). (4) Health-damaging behavior where the degree of choice of lifestyles is severely restricted. (5) Exposure to unhealthy, stressful living and working conditions. (6) Inadequate access to essential health and other public services. (7) Natural selection or health-related social mobility, involving the tendency for sick people to move down the social scale". Hertzman et al. advance six explanations: (1) Reverse causality (i.e., health causes social mobility). (2) Differential susceptibility (population categories have different abilities to resist environmental exposures). (3) Individual lifestyle. (4) Physical environment. (5) Social environment (and psychological response). (6) Differential access to/response to health care services. Some points are similar in both classifications, but their differences reflect that Whitehead is more concerned about social processes, while Hertzman et al.'s main focus is the circumstances which produce health.

Thus, most explanations found in the research field can be divided both as to how they conceptualise the social processes which are assumed to be involved, and how they treat the question of health determinants. Some remarks on these two aspects follow, before I outline the way I propose to divide the field.

3.1.2. Social processes: general remarks

A fundamental question for every explanation of social inequalities in health is how the social structures with their different social positions have emerged. The most fundamental "cause" of social inequalities in health is obviously that historical developments have released processes which have produced the existing social structures and their sets of more of less diverging social positions. Usually, this theme is not part of the explanations of social inequalities in health, because they take the existing social structures as *givens*.

When we start with the existing social structures, an important dividing line is whether explanations emphasise *social causation* or *social mobility* processes.

The basic tenet of social causation explanations is that social positions cause the health of the incumbents of the social position. Some type of cause-effect link from the circumstances typical of the social position to the typical health of the individuals in the position is assumed. Health inequalities are believed to arise because the circumstances of different social positions are so diverging that each of them develops a particular health profile, different from the health profiles of other social positions.

The explanations employing social causation processes often presuppose that people are, by and large, similar as to their innate potential for attaining health. On average, people in different social positions are believed to be equipped at the outset with more or less the same probabilities for health and ill health. Thus, the sources of social inequalities in health come from "outside" and are external to the members of the social positions.

Social causation explanations raise the question about *how* the circumstances of the social position act on the members of the position. How this happens can be accounted for in different ways. Explanations are divided as regards how they understand, in more detail, the characteristics of the social causation processes, and the main variants will be discussed in the following chapters.

While the key question asked by social causation explanations is "what circumstances are the individuals in particular social positions exposed to?", the key question of social mobility explanations is rather "what *individuals* are the different social positions exposed to?" The central tenet of social mobility

explanation is that health inequalities are generated because of how people stay in, move out of, or move into, social positions.

Terms such as social selection, health selection, health-related selection, health-selective mobility, health-related mobility, and even "natural" selection are often used when researchers discuss social mobility explanations. The word "selection" is frequently employed, implying that people are "selected" or "sorted" into particular social positions because of their health characteristics.

I will propose that social mobility explanations are a better heading for this group of social processes. Essentially, this explanation says that movements between social positions influence how social inequalities emerge. Social mobility may be health-related, so that health variations are associated with the types of mobility (or non-mobility) the individual experiences. However, even when social mobility is not health-related, mobility may redistribute people across social positions with consequences for the pattern of health inequalities.

When social mobility is associated with health, rather different processes may be the reason for this association. Health may in some way cause mobility, but "selection", and in particular "natural selection", is not always a suitable term for these processes. A person may choose a particular career because of health, employers may discriminate against ill people, or welfare state arrangements may encourage particular career choices. Health may be a main or a contributory cause in these processes. When this happens, it is not "natural". It is a social process (West 1991), involving the actions of agents (the person him/herself, families, teachers, employers) and structured by the functioning of institutions such as schools, firms, and welfare state organisations. It is based on social evaluations about how health and its different aspects (activity restrictions, physical strength, psychic balance, subjective well-being, etc.) *count*, and these evaluations are socially created.

Social mobility and health may also be statistically associated, without a causal connection. This is often referred to as *indirect selection* (Wilkinson 1986a: 4, West 1991, Dahl 1996). Indirect selection is health-related mobility where mobility is caused by other factors than health. Both when health is causally linked to mobility, and when it is statistically associated without any causality involved, mobility will redistribute health profiles among social positions and influence the distribution of health.

3.1.3. Health determinants: general remarks

Explanations are also divided as to how they understand the factors which produce health. Explanations of social inequalities in health can be distinguished as to what types of health determinants they employ.

We may define a health determinant as a bloc of factors or conditions which are presumed to have a general influence on people's health, their longevity, and their level of ill health. A health determinant is a main factor, or a main group of factors, which is presumed to influence the general level of health. In health inequalities research, it has also a more specific meaning. The health determinants are proposed to create, not primarily the average health of the population, but the *variation* in health between social positions.

Medical epidemiology is also concerned with the explanation of ill health, but health determinants are not immediately identical with the factors employed by medical epidemiology. Medical epidemiology is dominated by "the doctrine of specific etiology" (Dubos 1995:6, Conrad 1999). It translates ill health into specific conditions (cf. section 2.4.2). In order to explain how such specific conditions emerge, medical epidemiology focusses on specific *risk factors* – smoking, shift work, genes, lack of vitamins, etc. The objective is to try, often by means of multivariate methods, to reveal the significance of each risk factor for the incidence of specific diseases (cf. Popay et al. 1998: 627pp).

However, in health inequalities research, a health determinant is often supposed to have generalised health effects. This idea corresponds to the observations that disadvantaged social positions usually suffer from bad health in a general, multifaceted way. Earlier, it was argued (section 2.4.4) that health displays unidimensional characteristics at the aggregate level. Although there are important exceptions, health seems to vary between social positions in more or less the same way, whatever health indicator we focus on: premature deaths, injuries, chronic disease, or psychological problems.

This suggests that there is a broad set of conditions which acts on the social position in a way which has multiple, unspecific consequences for health, and which therefore produces ill health, or good health, *generally*. This idea has been advanced in terms of the concept of *general susceptibility*. Syme and Berkman (1976, also Locker 1997:20) argue that the reason for the higher frequency of health problems, of a most varied kind, in less advantageous social positions, must be found in "something about the life in the lower classes (that) increases vulnerability to illness in general". General susceptibility implies "generalised compromises of disease defence systems". Hertzman et al. (1994:80) argue similarly, using the terms *generalised vulnerability*.

The general susceptibility/vulnerability idea suggests to rearrange how we think of the relation between specific diseases and ill health. Diseases may be conceived of as "pathways" or "mechanisms" (Evans 1994:7) which reveal that health has been impaired. Specific health problems are seen not as the essential phenomenon, but rather as the concrete manifestations of a deteriorated health.

It may be said that while medical epidemiology supposes that disease is the cause of ill health, the general susceptibility/vulnerability view reverses this notion and supposes that disease is a *consequence* of ill health.

The concept of *general susceptibility* has a low status in established medicine, compared with disease-specific etiology (Blane et al. 1993b: 76). It is obviously incongruent with the efforts to understand and cure the specific diseases of individuals which doctors are confronted with. However, also medical epidemiology has produced evidence which points towards the general susceptibility view. It is not seldom found that risk factors such as deficit nutrition or stress have very wide negative consequences and can be associated with quite different types of disease.

When researchers study social inequalities in health, their focus may of course be social variations in specific diseases. More often, however, they are oriented towards understanding social variations in health and ill health in a more general way, and they often employ composite and all-round health indicators such as all-cause mortality, longstanding illness which limits working capacity, self-perceived general health status, or health expectancy. That there are health determinants which influence the overall health of social positions and their general susceptibility to ill health is therefore often the implicit understanding of many of the explanations which are proposed.

Thus, the explanations of social inequalities in health are divided as regards what types of health determinants they primarily rely on, and various propositions about such health determinants will be discussed in the following.

This must however be understood in a particular way. When a certain health determinant is claimed to be responsible for social variations in health, this means that the focus is on how this determinant *varies* between social positions. As Rose (1985) has pointed out, the determinants of the *average* level of health in a population are not necessarily the same as the determinants of the *variations* in health within the same population. If, for instance, it is proposed that the psychosocial environments are the main determinant of health inequalities, this does not necessarily imply that physical environments are irrelevant for population health. The physical environments in a society may be the major reason that average longevity in the whole population is, say, 75 years, and not 65 or 55 years. But if the physical environments do not vary significantly between social positions, the importance of this health determinant is primarily to create the overall level of health, while other determinants are more relevant as regards the social *variations* in health.

Thus, when explanations point to the physical environments, or the psychosocial environments, or some other conditions as a main health

determinant, the implication is that those health determinants which are not emphasised, do not vary in any important way within the population, and therefore they are of less importance for the health variations in the population.

3.2. Overview of the explanations

The following chapters try to *cover* the research field, not with too many details, of course, but with the aim of making a comprehensive review of all the major alternatives the research field has proposed as ways of accounting for how social inequalities in health emerge. As said above, the complexity of the field implies that the dividing lines may be drawn in different ways. The way I propose to do it, is therefore not the only feasible one.

First, I will address the so-called *artefactual* explanation (Chapter 4). Often, it is not included when explanations of health inequalities are listed, but it was highlighted by the Black Report as one major alternative. It is special in the sense that it primarily addresses how the *picture* researchers construct of health inequalities depends on the ways researchers classify, measure and interpret their data. Thus, the artefactual explanation explains health inequalities in terms of the research process. One of the five papers included in this thesis addresses this explanation, and this is also a reason for discussing it separately.

Thereafter, I will treat what we may call substantive explanations, as they are oriented to the question about how health inequalities are produced, not primarily in terms of processes internal to the research process, but in terms of how "real" social processes and health determinants interact in society.

As pointed out before, whether we focus on health determinants or on social processes may give rise to different ways of dividing the field. The choices I have made are the following.

There is a major dividing line between explanations according to whether they assume social causation or social mobility as the main social process. This dividing line is applied here, so that Chapters 5 – 7 address social causation explanations, while Chapter 8 treats explanations which have social mobility as the main, or at least as an important, social process.

The social causation explanations are however divided in several ways, both according to the specific ways they conceptualise the social causation process, but also according to the health determinants they rely on. There are alternatives which employ one special type of social causation process, but differ as to the health determinants they employ, and on the other hand we have social causation explanations which focus on one special health determinant, but differ as to the type of social causation they emphasise.

In order to clarify these distinctions, I divide social causation explanations into three, the *materialist/structural explanation* (Chapter 5), the *behavioural/ lifestyles explanation* (Chapter 6), and the *psychosocial explanation* (Chapter 7). Although there are many subvariants within each of these approaches, it may nevertheless be argued that each of them constitutes a relatively coherent alternative, different from the others.

In the materialist/structural explanation, social causation is usually understood in terms of structural determination. As regards health determinants, this explanation emphasises those physical environments which are constituted by unhealthy working conditions and the restrictions on material levels of living imposed by a low level of socioeconomic development.

The behavioural/lifestyles explanation is generally characterised by a formulation of the social causation process which allows for a larger role of individual agency than the materialist/structural explanation. As regards health determinants, it is also oriented towards the physical environments, but it emphasises primarily the physical environments which arise from people's behavioural and consumption patterns.

The psychosocial explanation employs various ways of conceptualising social causation, partly coinciding with, but also at variance with, the ways the two former explanations understand social causation. The distinguishing feature of this explanation is however that it emphasises the psychosocial environments as the main health determinant.

Finally, chapter 8 addresses *social mobility* explanations. The chapter addresses various ways of understanding how social mobility contributes in generating social inequalities in health. Social mobility explanations often assume that people's health potential, established early in life, is a main health determinant. But there are also variants which take as their starting point that health differs between people, for whatever reason, and this has consequences for people's life trajectories. Thus, this explanation is more unspecific as regards the health determinants than the former types of explanations.

The following Chapters 4 - 8 will include some remarks on the history of the explanations, describe main features, main empirical themes, and important subvariants, and point to some problematic and disputed aspects within them. I will also show that some of them, when confronted with difficulties, often try to include issues which "belong" to other explanations. Thus, within each explanation there are tendencies which point towards an integration of the explanations, a theme which will be further discussed in the concluding Chapter 10.

4. The artefactual explanation

4.1. Main characteristics

4.1.1. Background

The question whether "artefacts" were responsible for findings of health inequalities was raised by the Black Report. According to this report, the artefactual explanation suggests that "both health and class are artificial variables", often measured arbitrarily. Following from this, the picture researchers make of the health inequalities in society may also be an artificial product, which appears mainly because of how researchers choose to arrange and present their data. The implication was that findings of health inequalities should not be taken too seriously, as they could be products of the methods chosen. For the same reasons, inferences about the relationship between health and social class would be uncertain (see Macintyre 1997: 727, Bury 1997: 61).

The Black Report rejected the artefactual explanation. The debate that followed was stimulated by the political reactions to the Black Report. The Conservative government at that time (1980) neglected the report and restricted its publication to 260 copies (Dahl 1994b: 1). The message of the Black Report was that the continuing existence and even *worsening* of health inequalities between occupational classes were alarming. The proponents of artefactual interpretations argued that the findings of the Black Report called for no urgent action. The debate was "vociferous and acrimonious" (Macintyre 1997: 732).

4.1.2. The issues involved

The issues involved in the debate about the artefactual explanation are quite many. A common theme is however that focus is directed at the *research* process. One looks at the activities of researchers and examines to what extent the results are reflections of the ways researchers have proceeded. Thus, the explanation of health inequalities is not primarily sought in processes operating in society "out there", but rather in processes internal to research.

One issue is therefore *methodological*. An aspect of this explanation is a theme pertaining to all social research, that findings are not indifferent to the choices, made by researchers, as regards definitions, measurements, and ways of arranging and analysing data. Artefacts point to how research can construct different images of the size of health inequalities, their developments, and their causes, depending on how data are obtained and analysed.

However, when this familiar methodological issue is presented in terms of "artefacts", a wider problematic is introduced. The question is now not only that the phenomenon – social inequalities in health – may be described and interpreted in different ways. The question is also raised about the "reality" of the phenomenon itself. Could it be that it has no substance, but is rather an *artificial* object, produced by researchers and the research process?

No researcher has, to my knowledge, simply denied the existence of at least *some* social inequalities in health. When the question about methods is reformulated as a question about artefacts, the implication is however usually that the images of these inequalities, produced by researchers, are artificially magnified, the size of inequalities is regularly exaggerated, and the reasons for social concern are unduly overstated. Thus, it is suggested that the results produced by researchers are biased in a special way, making health inequalities a much larger social problem than they "really" are.

However, researchers have pointed out that artefacts, understood as biased results because of inappropriate methods, can also result in an *underestimation* of the size of inequalities, and not only in exaggerating them (Bloor et al. 1987, Dahl 1994b: 6). The research process may create or exaggerate the problem of health inequalities, but it may also make the phenomenon diminish or even disappear, "artificially". A recent analysis of the consequences of methodological choices for the resulting magnitude of inequalities shows this (Kunst et al. 1998). This study focussed on the numerator/denominator problem (see below), on the inclusion or exclusion of economically inactive people in the analysed sample, and on different social class schemes. The conclusion was that methodological choices were important, but whether they meant overrating the inequalities or the opposite was often hard to establish.

In the following, I will survey some main issues involved in the artefactual explanation. I will focus on questions about measurements: how are the images, and in particular the *size*, of social inequalities in health influenced by the way researchers have measured health and social positions? Moreover, what issues are involved when we measure inequality? Lastly, the interpretation of *trends* will be discussed.

4.2. Questions about measurements

Key terms in every study within this field will always be "health", "social positions", and "inequality". How to measure these concepts is therefore a recurrent topic. In the following, some of the typical questions which are raised as regards these measurements will be outlined. The focus is how

various processes may interfere in the measurements in a way which may bias (diminish or magnify) the picture of health inequalities.

4.2.1. Class biases in measurements of health?

How to measure death in itself constitutes no problem in health inequalities research. The ambiguities concerning when a person may be considered dead, relevant for high-tech hospitals and transplantation issues, will hardly have any consequences for how deaths in general are registered among individuals in different social positions.

When the issue is *causes* of death (cause-specific mortality), however, the question is regularly raised whether these causes are validly registered, and in particular whether they are registered in the same way, whatever social position the deceased person belonged to.

In Britain, it has for instance been observed that deaths in the middle class, more often than in the working class, are registered with more than *one* cause of death. It has been supposed that physicians somehow believe that middle class deaths are more hard to comprehend and therefore "in need of more explanation than working-class deaths" (Nettleton 1995: 162). How this influences the observed distribution of causes of deaths between social classes is however unclear.

Bloor et al.'s study (1987) suggests that class patterns in how physicians report causes of death have given rise to an artificially produced issue in the health inequality field. A widespread opinion has been that the social distribution of deaths from coronary heart disease (CHD) has been altered (Marmot et al. 1978, Davey Smith 1997: 248pp, see also Kruger et al. 1995, Aase and Storm-Furru 1996: 37). Many researchers believe that previously, for instance during the 1940s and 1950s, CHD was "the rich man's disease". However, in recent decades data indicate that CHD shows the familiar inverse relationship with social status. Bloor et al. suggest that this apparent change could be explained by changes in diagnostication. Earlier, physicians may have been more superficial when examining working class deaths. Therefore, the frequency of CHD in the working class may have been underestimated. Now, physicians are more specific and thorough, also as regards working class health problems. The apparent change in the social distribution of CHD deaths, it is suggested, is therefore an artefact, produced by changes over time in class biases in how the causes of deaths are registered.

These examples show how physicians' verdicts about cause of death may interact with physicians' knowledge about the social position of the deceased. Thus, the data which lie behind findings about patterns of social inequalities in

deaths from specific causes may be distorted. Similar processes may occur when morbidity is measured. When research is based on morbidity information from physicians, the possibility exists that physicians' diagnoses are not made independent of the social position of the patient. It has for instance been suggested that women with appearances and behaviours which indicate low status, easier are given vague diagnoses such as fibrositis or fibromyalgia (Christensen 1996).

The use of medical and professional judgement in studies of health inequalities is therefore no guarantee against biased health measurements, and various studies suggest that professional verdicts as regards causes of death and disease diagnoses are not given independent of the patient's social position.

Studies of social inequalities in health will often measure health in terms of *self-reports* in interviews and questionnaires. The methods used (personal interviews or self-administered questionnaires, for instance) give different results (Moum 1998), but whether this influences the class patterns is not well known. Generally, if different social positions have different ways of perceiving and evaluating health complaints, and perhaps also have different ways of using professional help for diagnosing symptoms, self-reports may give rise to distorted estimations of the health differences.

Thürmer, for instance, found more *un*diagnosed symptoms of heart disease among workers than among middle class men, and suggests that "the health care system favours the detection and treatment of asymptomatic risk in high status men more than low status men" (1993: 94). It may be that Norwegian physicians are less scrupulous when examining lower class men, so that symptoms are not discovered, but it may also be that lower class men do not seek professional help for symptoms that the middle/upper classes more often consider a cause for a visit to the doctor. To a large degree, self-reporting of diseases is dependent on the diagnoses communicated by doctors to the person. Thus, if workers, as Thürmer's study suggests, either are more reluctant to visit doctors, or when visiting doctors are examined more superficially than middle class patients, the result will easily be that self-reports underestimate the size of the social class health differentials.

Some studies have suggested that the size of the class differentials in symptoms and morbidity may be overrated because of a higher tendency in lower classes to complain about trivial health problems. This issue is treated in one of the papers included in this thesis (Elstad 1996b). Studies have usually not found indications that the higher morbidity among women compared to men is due to lower thresholds for reporting pain among women (Bendelow 1993, Annandale 1998: 147, Macintyre et al. 1999). – How cultural

differences, linked to ethnic or national background, interfere in the reporting of ill health has been widely recognised (Zola 1966, Koopman et al. 1984). This suggests that self-reports are not an immediately reliable way of describing ethnic health differences.

4.2.2. Problems in measuring social positions

To classify people according to their social position in the gender and age structure is easy. Other social structures do regularly present many problems related to how to define the social positions in the social structure and the detailed procedures for classifying persons into them. *Ethnicity*, for instance, is often difficult to measure (Nettleton 1995: 185-186). How people are classified in the ethnic structure depends on whether one focusses on nationality, citizenship, appearance, language, or cultural belonging. Anthropological studies in Norway have for instance tried to establish "who are the Lapps?", but the answer is vague (e.g. Hovland 1999). *Marital status* can also be ambiguous (Mastekaasa 1993a: 8-9). Marriage can involve formal arrangements, cohabitation, monogamy, and sexual relations, and differences in classifications occur according to what aspect is made the central one for determining the marital status of the individual. To what extent choices made when classifying ethnicity and marital status have consequences for how the health inequalities are pictured, is however unclear.

How *socioeconomic position* is measured has attracted particular interest within the research field. Occupational categories are usually employed when the social class or stratification structure is investigated, but income and educational level are also regularly used. Whether occupational categories show more marked health gradients between socioeconomic positions than income or education has been examined (for instance Dahl 1994a), but the conclusion is not clear. There are many ways of classifying occupations, and many are used in health inequality research (Elstad 1995b). Important questions concern how the social positions within this structure should be delineated, and to what extent they constitute a hierarchy or are only nominal types with no hierarchical ordering. Often it seems, however, that variations in the definitions of occupational classes do not lead to drastically different views of the size of socioeconomic health differentials (Bartley et al. 1996, Wohlfarth 1997, Bartley et al. 1999). One question is how *homogeneous* the occupational categories are, as regards the factors which are assumed to influence health, such as working conditions, income, and prestige (Prandy 1999). Whatever principles employed for constructing occupational categories, the result seems always to be that health varies widely within each occupational category.

The question about women's socioeconomic position – should it be measured in terms of own occupation, or in terms of her husband's, or in terms of some composite measure for the status of the household – has been another issue. Dahl's (1991) study suggests that socioeconomic health differentials are larger when women are classified according to husband's occupation than when classified according to own occupation. Krieger et al. (1999) find however that a composite, gender-neutral way of classifying households produces the largest differentials between socioeconomic positions among women. These findings are not necessarily conflicting – it may be that circumstances are different in Norway (Dahl's study) and the U.S. (Krieger et al.'s study) – but they nevertheless suggest that, at least as regards women, the way research proceeds for classifying their socioeconomic location has distinct consequences for how we perceive the problem of health inequalities among women.

Further questions about the measurement of health inequalities between social positions in the social class or stratification structure are raised by mobility, particularly *out* of the labour market. Exclusions from paid work because of health problems occur more often among low status occupations (Dahl and Birkelund 1999). The magnitude of health inequalities is influenced by the "healthy worker effect", i.e., the tendency that healthy workers are more fit to survive in occupations which require physical strength and capacity (Dahl 1993a, 1993b). Therefore, health inequalities are usually less marked when using current occupation than when including also non-employed in the analysed sample and classifying them according to their original occupation. Measurements which indicate a more long-term location in the social hierarchy, such as educational level or life-time income, are therefore often preferred in order to obtain more "true" estimates of the size of health inequalities in the stratification structure (Robert and House 2000: 116).

Another issue particularly related to the social mobility explanation for health inequalities concerns how vertical mobility is measured. The social mobility explanation will usually involve a comparison of the individual's social position at one time point with his/her social position at an earlier time point, for instance earlier in the respondent's occupational career, or the social background of the respondent. The understanding of what constitutes social mobility is dependent on how occupations are measured, and how occupational structures in different historical epochs are adjusted to each other. Wilkinson has suggested that when intergenerational mobility is studied, one should be aware that information about own current occupation often will be much more precise than information about earlier, or father's, occupation (Wilkinson 1986a: 9). Errors in classifications of mobility types may perhaps produce artificial inferences about the association between health and mobility.

The so-called numerator/denominator problem has attracted considerable attention. This problem concerns how numerators and denominators in the equation that calculates mortality rates of different social classes are estimated. A particular problem in the British setting has been the practice of estimating occupational class mortality differentials by "dividing the number of deaths in occupations stated on the death certificates by the number of people in each occupation as recorded at census" (Wilkinson 1986a: 3). Thus, the reports of occupations refer to different time points, making it uncertain whether the number of deaths reported in a particular social class can immediately be related to the number of people at risk as indicated by the census information.

Moreover, some British studies indicate that the occupation of the deceased registered at the death certificate is reported by mourners who sometimes want to "promote the dead" in order to make him or her more worthy (Wilkinson 1986a: 3, Annandale 1998: 112). If so, the number of middle/upper class deaths are artificially elevated, resulting in an underestimation of the "true" differentials in occupational mortality.

The numerator/denominator problem was in particular a part of the British artefactual debate, but, according to Wilkinson (1986a:3), when longitudinal studies were established in Britain which were able to use occupational measurements at the same time points, the mortality differences between occupational classes turned out to be very similar to the results which were suspected of being biased because of the numerator/denominator problem.

In Norway, the numerator/denominator problem is somewhat different, as the occupation of the deceased is usually found by linking mortality statistics to census information. However, here the problem is that there is usually no certainty whether the deceased held the same occupation at the time of death as he/she had at the time of the census.

4.4.3. Measurements of inequality

The research field addresses health inequalities between social positions, but inequality is an ambiguous concept. It may include various phenomena, such as injustice, absolute disadvantage, relative disadvantage, inequality in opportunities, or inequality in outcome.

When one tries to establish the magnitude of health inequalities in a social structure, there is a distinction between the size of the health differences between social positions, and to what extent the total distribution of health in the whole social structure deviates from equality (Dahl 1994b: 25-56, Mackenbach and Kunst 1997, Manor et al. 1997, Hellevik 1997, 2000, see also Marshall and Swift 1999). Inequality can be conceptualised as the *distance* between the social

positions in average health, i.e., as the "effect" on health of belonging to one social position instead of another. Inequality can also be conceived of as "*unrepresentativity*", i.e., a distribution of ill health which is not corresponding to the distribution of social positions (Hellevik 1997: 377).

Distances between social positions are for instance measured by differences in mean values, percentages, or odds ratios. This way of seeing inequality does not take into account the proportion of the population which belong to different social positions. Differences in mean values, percentages, and odds ratios between, say, the working class and the middle class, will be the same no matter whether the working class is smaller or larger than the middle class.

Unrepresentativity can be measured by for instance gini coefficients (Hellevik 1997) or concentration indices (Wagstaff et al. 1991, Dahl and Elstad 1999). Such measurements are sensitive to the *relative* proportion of each social position in the population, so that not only the distance (for instance large health differences between workers and white collar occupations) but also the relative number of workers and white collar are taken into account.

These two ways of conceptualising inequality are different. One may say that the first one points to questions of justice, and indicates the disadvantages of belonging to one social position instead of another. How the likelihood of being afflicted with ill health differs for the *individuals* who belong to different social positions is measured. The second one does not measure average individual disadvantages, but makes an index of all the individual disadvantages/advantages in the social structure. Thus, it may be said that the second one indicates the amount of resources which would have to be invested in order to eliminate inequality.

These two approaches may sometimes give quite diverging impressions of how large the inequalities are, especially if there is one social position with few members which has a particularly high level of ill health, while the average health of the other social positions is relatively similar. When we focus on the health disadvantages of the "extreme" social position, inequality will appear quite drastic, but when we focus on how much the total distribution of health deviates from representativity, the deviation will be quite small, because the deviating social position constitutes but a small proportion of the whole population.

Another aspect when measuring inequality in terms of disadvantages or effects concerns *absolute* versus *relative* differences. The difference in mortality rates (for instance number of deaths per 100.000 persons, often age adjusted) is an absolute figure, while the difference in standardised mortality ratios is a

relative measure based on mortality rates as a percentage of, for instance, the national average. Moreover, differences in mortality between social positions can be measured in terms of PYLL (potential years of life lost), which takes an expected longevity of, for instance, 75 years as the norm and calculates how many years of life each social class looses because of deaths occurring before the age of 75 years. PYLL is especially sensitive to the harms of deaths among young people and gives sometimes larger differences in mortality between occupational classes than standardised mortality ratios (Dahl 1994b: 6, 37).

4.4.4. The interpretation of trends

Thus, inequalities in health in a social structure can be measured in quite a number of ways – as absolute or relative differences between social positions, as distances between social positions or deviations from representativity, inequalities in mortality may be represented by mortality ratios or by PYLL, etc. These differing ways of measuring inequality may give somewhat diverging impressions of how large and socially important the health inequalities are. Why this happens is mostly easily understood, and while one may debate what way of conceptualising and measuring inequality should be the preferred one, these ways are seldom drastically conflicting. If health inequalities are present in the social structure, all the different measurements will usually indicate this, although they do it in somewhat different ways.

However, when one asks not only whether social inequalities in health are present, but in addition wants to know how they develop over time, the choice of inequality measurements will often be a consequential issue. In the debate about the artefactual explanation, a main theme has been the interpretation of trends, i.e., are health inequalities becoming larger or smaller? This involves the comparison between the inequalities in a particular social structure at different time points. The interpretation of trends is often complicated because different ways of measuring inequality may indicate different types of trends.

How to judge the development of health differences in a period when health is improving is thus dependent on the choice of inequality measurements. A common pattern will often be that when mortality is decreasing in the population, absolute differences will decline while relative differences remain or even increase (Bury 1997: 56). Comparing mortality in the two Norwegian counties *Finnmark* and *Sogn og Fjordane*, one finds for instance that the difference in standardised mortality ratios has increased since the 1960s (117 and 87, respectively, in 1964-67 and 120 and 88 in 1991-95), although absolute differences in age- and sex-adjusted yearly number of deaths per 100.000 have decreased (Statistics Norway 1969: table 4, Statistics

Norway 1998b: table 24). Thus, how we interpret the trend depends on whether relative or absolute measurements are being used.

Similarly, when one measures inequality as distances between two social positions, whether inequalities appear to become smaller or larger depends on the way we measure this distance. Say, for instance, that 30 per cent of the working class and 15 per cent of the middle class have ill health at the first time point, and at the next time point the corresponding figures are 35 per cent and 20 per cent. The differences in proportions with ill health are stable, but if the health differences are measured as odds ratios, they appear to decline (from about 2.44 to about 2.15).

Thus, the picture of trends is often dependent on what type of inequality measurement one employs, especially when the average level of health is changing in the period under study. This will be even more complicated when both the average level of health and the social structure change simultaneously. If the health disadvantages of the working class increase, compared to the middle class, during a period when the working class is becoming smaller, inequality measurements which focus on the effects of being a member of the working class will usually indicate increasing health differences. Measurements which take into account that the disadvantaged sections of the population are shrinking (such as the gini coefficients or the concentration indices) will however often suggest that inequality, in terms of unrepresentativity, has decreased (Hellevik 1997).

The interpretation of trends was in particular in focus in the debate about artefactual explanations in Britain. The Black Report observed that the mortality differentials between the "extreme" occupational categories (Class I versus Class VII) had remained relatively constant in absolute terms, but increased in relative terms, from the 1950s to the 1970s. One theme was whether this indicated that mortality differences were increasing or not, and the answer to this question may depend, as we have seen, on the types of inequality measurements which are used. A further question was whether the trends in mortality should be understood as a result of widening differentials in the general social and material circumstances of the different social classes, i.e., as generated by social causation processes, or was a product of "health selection". The former view was advocated by Wilkinson (1986a, 1986b, 1987), while Illsley (1986, 1987) interpreted the trends mainly as consequences of social mobility processes. Thus, the debate about artefacts, which primarily focussed on the research process, moved towards discussions about how the "real" processes in society could generate health inequalities. The explanations which have them as their main themes will be treated in the next chapters.

5. The materialist/structural explanation

5.1. The basic model

The materialist/structural explanation has perhaps been the dominant approach for explaining social inequalities in health since population statisticians and social scientists started examining these questions. First, I will outline what could be termed the *basic model*. Later, I will show some of the difficulties which have confronted this model in recent decades. These difficulties have, on the one hand, suggested reformulations of the basic model. They have also inspired the formulation of explanatory alternatives which constitute a break with the materialist/structural explanation.

Social causation as structural determination and the physical environments, defined in a special way, are the two ground pillars of the basic model.

5.1.1. Structural determination

Societal developments, viewed from a macro perspective, is usually the background for the basic model. Through history, an evolution of economic, political and social institutions occurs, with corresponding major social structures and their social positions. The main positions in the socioeconomic structure, the marital structure, the regional structure, etc., are confronted with significantly differing physical environments, and the variations in the physical environments are seen as the major cause of the health inequalities.

A marxist view of the historical evolution of capitalism (e.g. Wright 1985, Wood 1986) is often the background for this model. However, also other "grand theories" can function as the basis for this explanation. The class structure as envisaged by Weber or Dahrendorf (Bottomore 1970), Durkheim's analysis of the division of labour in society (Durkheim 1964), the functionalist view of the necessity of differentiated rewards to different positions in society (Ritzer 1993: 68-69), and conceptualisations of the post-industrial social structure have in common with the marxist view that grand social forces lead to social inequalities which are characterised by different physical environments surrounding the different social positions.

The incumbents of each social position are therefore confronted with structurally determined conditions. People are *exposed* to the characteristics of the social positions they belong to. People are not choosing their circumstances. Rather, they are subjected to the conditions of the social position they occupy.

That people are creating their way of life, or have their destinies in their own hand, or that it is generally possible to move out of their current social positions, are not part of this basic model. The conditions they are confronted with, are created for them by macro-social processes largely outside their influence. When they act, their actions are best understood as reactions: structurally determined behavioural types and ways of living which reflect the compelling constraints emanating from the social structures. They are located in social positions, with typical material circumstances, working conditions, and standards of living. They cannot – in the typical case – free themselves from these externally determined circumstances. Their lot is structurally determined. Even stronger, they can be viewed as *victims* of the social structure. *Structural victimisation* is an apt term of the social processes this model operates with. The exposures and constraints typical of each position are unavoidable, and these exposures and constraints vary significantly between social positions and thereby produce social inequalities in health.

5.1.2. Health determinants: the physical environments

Thus, people are exposed to structurally determined circumstances, but *what* circumstances are central for health? The health determinants of this model are primarily seen as the *physical* environments. Underlying this basic model is an understanding of health as a product of how human bodies, as physical and "natural" entities, interact with the physical circumstances.

The relevant physical environments are specified in a special way. Focus is on the work life, the material level of living, and the physical characteristics of residential areas. Here, people accumulate exposures to substances and material "things". These environments, which are unavoidable parts of their surroundings, are regarded as the most significant determinants of their health.

At work, people handle physical, chemical and biological substances, directly or through the air, which may produce cancers, respiratory diseases, skin disorders, and other somatic diseases. Noise, heavy lifting, a high working pace, monotonous movements, long working hours, and shift work lead to bodily strains and exhaustion. The organisation of the working place may constitute unavoidable physical risks and probabilities of work accidents. Next, people's income from work determines their material conditions. The level of nutrition, and the frequency of overcrowded, damp, cold, or unhygienic dwellings, determine the level of malnutrition and the spread of germ-related diseases. The ability to buy decent clothes, and the possibilities for rest and holidays, are other parts of the material standard of living which enhance or jeopardise health. The residential area, with its level of air pollution, its

provision of clean or polluted water, its organisation of sewage disposal, and the availability of safe playing grounds for children, are further components of the physical circumstances which determine health (Blane et al. 1997).

The health determinants of this model is conceptualised in a way which clearly is related to the biomedical notions of disease etiology. The *host* is surrounded by physical *environments* with pathogenic *agents*, and from this interaction, the level of health is produced. Cause-specific mortality and disease-specific health disorders are often used as indicators of health in this model, alongside more general measurements of overall health. The *health services* are often considered an important part of the physical environment. Vaccines, treatment, and therapies are included as significant physical determinants of health. The organisation and payment systems of the health services make the availability, affordability, and use of these services differ between social positions.

5.1.3. The classic study

The prototype of the basic model can be found in Friedrich Engels' *The Condition of the Working Class in England* (1969 [1845]). Engels starts his account with the growth of capitalist production and the parallel growth of the industrial proletariat since the latter half of the 18th century. The steam engine and machinery for working cotton were invented, the factory system of organising production was implemented, the working population in the industrial towns increased rapidly, and in general the industrial revolution "altered the whole civil society" (1969: 37).

These societal developments created and determined the conditions of the working class. Drawing on official inquiries, reports from factory inspectors, and own observations, Engels analysed the lot of the new-born proletariat. Poverty abounded, unhygienic and overcrowded dwellings were the rule, and malnutrition occurred frequently – the workingmen's meal "consists almost exclusively of potatoes, with perhaps oatmeal porridge, rarely milk, and scarcely ever meat" (1969: 169). The working conditions inside the factories were hard and unhealthy, and the employment of children had particularly severe consequences. Children were employed "rarely (at the age) of five years, often of six, very often of seven, usually of eight to nine years ... the working day often lasted fourteen to sixteen hours" (p.179). Engels asked polemically: "How is it possible, under such conditions, for the lower class to be healthy and long lived? What else can be expected than an excessive mortality, an unbroken series of epidemics, a progressive deterioration in the physique of the working population?" (p.129) Not only diseases, but also

crippled and deformed bodies occurred frequently among the working class. Miners "who begin this work in early youth are far from reaching the stature of women". Usually, "they age prematurely and become unfit for work between the thirty-fifth and forty-fifth years" (p.268).

5.1.4. The tradition of social medicine

Engels' work is of course unique in many ways, but it is also part of the *social medicine* tradition as regards population health and health inequalities. It originated long before Engels. Johann Peter Frank (1745-1821), the author of a nine-volume *System einer vollständigen medizinischen Polizei* (Kock 1955: 104), was one of its founding fathers. Ill health was seen as a product of economic deprivation and health-damaging working conditions. The cures were reorganisation of the hygienic environment, the regulation of working life, the improvement of dwellings, the eradication of poverty, and access to health services. Social reform was part of the programme for social medicine, and social medicine meant political involvement. "Medicine is a social science, and politics is just medicine at large" was the credo of another main figure in 19th century social medicine, Rudolf Virchow (1821-1902).

Social medicine was not seldom allied to the workers' movement, and this affinity was not always wished welcome. The Norwegian Director-General of health services Karl Evang remarked that American doctors sometimes were unable to distinguish between social medicine and *socialist* medicine (Evang 1974: 44). But public efforts and governmental action in order to improve the health of the working class and the poorer sections of society were not only the policies of the left. Also conservative could adhere to public health programmes, if not for other reasons, at least because the frequent unfitness of young workers "might impede economic and military success" (Kuh and Davey Smith 1997: 16), and this was alarming in the period of capitalist industrialisation and imperialism.

5.1.5. Working conditions and material deprivation

During the 20th century, the focus of the materialist/structural explanation has continued to be working conditions and material deprivation.

Occupational medicine is one medical branch of this tendency, concentrating on health-damaging air, material substances, and physical characteristics of the work situation, and the consequences of this for the health of workers (Blane et al. 1997). Also many medical sociology studies emphasise the overall importance of variations in the working conditions for social class inequalities in health. One study based on the Norwegian *Health Survey 1975*

finds for instance that "the major part of the class differences (in chronic diseases emerging after the age of 25) can be accounted for by the variations in working conditions" (Elstad 1981: 181). Lundberg, using Swedish data, concludes similarly that "physical working conditions, together with economic hardship during childhood and to some degree health-related behaviours, seem to be the most important factors" for socioeconomic differentials in health (Lundberg 1991). Studies of working life is not restricted to paid employment. Bartley et al. (1992) consider for instance how working load and other material circumstances of home-working women contribute to their health problems.

The faces of poverty has changed, but the correlation between health and economic resources and material standards of living is a recurrent finding. Dahl (1994a), for instance, reports how income has an independent effect on health variations, after controlling for education and occupational status – possibly implying the direct effects of different material levels of living. A recent trend is the application of more direct indicators of material resources. Instead of income, so-called *asset-based measures* (ownership of car, ownership of house, etc.) are believed to constitute a better approach to measure economic resources. Using such indicators one finds "consistent and marked differentials in morbidity, mortality and risk factors" (Macintyre 1997: 732).

Another typical approach is the study of health variations between local areas described in terms of dwelling characteristics, average income, and other indicators of the physical circumstances. Although warning against the possibility of committing the ecological fallacy, Abbott and Sapsford, in their study of Plymouth (England), conclude that "it seems ... very likely ... that people's material deprivation is indeed one of the factors predictive of their poor health" (1994: 258). A number of similar reports, showing how local communities' material deprivation is connected to their level of health, have been published (e.g. Blane et al. 1996a, Curtis and Jones 1998).

5.2. Debates: the relevance of the basic model in affluent societies

5.2.1. The threshold notion of the basic model

Usually, the basic model of the materialist/structural explanation operates with a particular notion of the health-damaging process. Health maintenance is supposed to require that the physical environments have certain *absolute levels*: sufficient unpolluted air, sufficient nutrition, sufficient rest, sufficient dwelling standards, and sufficient access to health services. When the physical

environments are deviating negatively from these levels, considered necessary for the normal development and functioning of the body, ill health occurs.

The implication is that if these absolute levels are attained, the causes of health inequalities are eliminated. If people, by and large, whatever their location in the social structure, have enough food, sufficiently hygienic and healthy dwellings, working conditions which do not exhaust and overload the body, and access to necessary health services, the reason why health differences between social positions emerge could be assumed to have disappeared.

Social developments during the 20th century have put the basic model under strain. The arguments are of course not that poverty has been eliminated, or that working conditions no longer contain health-damaging elements. But overall, the physical environments have changed markedly for all social classes and all social positions. Compared to some 50 or 100 years earlier, during the 1970s, 1980s and 1990s, the majority of the working class were employed in work places with less health risks. Dwellings are better, overcrowding is hardly a widespread phenomenon, the material standard of living is markedly improved, and access to health services is not severely restricted. A telling example can be found in Norwegian surveys. Among cohorts born during the first decades of the 20th century, nearly 60 per cent reported economic problems in their childhood family, and 22 per cent reported lack of food, at least now and then. The corresponding figures for those born during the 1960s were 10 per cent (economic problems) and 3 per cent (lack of food) (Elstad 1999b). Such figures suggest that in Norway and similar countries, large sections of the population were *below* the absolute levels required for the maintenance of health during the first parts of the 20th century, while the large majority are placed *above* such levels during the latter third of the 20th century.

5.2.2. Continuing differences, the challenge of the gradient

What has happened with population health in general, and with the social inequalities in health in particular, during these transformations?

Mortality has changed dramatically. In Norway, average life expectancy for the newborn increased by about 30 years from the 1860s to the 1990s (Statistics Norway 1998a: table 60). Infant and childhood mortality has been drastically reduced. This is connected to the *epidemiological transition* (Wilkinson 1994, 1996: 29pp). Nowadays, deaths occur primarily among the middle aged and even more among the older sections of the population. Contagious diseases dominated the health panorama and cause-specific mortality statistics during the 19th century. During the last part of the 20th

century, heart disease, cancer, and accidents are the main causes of deaths. The illness panorama is dominated by longterm, not immediately lethal, chronic diseases, and by various symptoms and psychologically related conditions.

The health of every major social position in society mirrors, more or less, this general development. The major picture is however that, although there has been a continuing transformation of the average levels of health in the population, the differences between the major social positions within most social structures have undergone surprisingly small changes. When the average level is changing, the measurement of differences in terms of absolute or relative indicators can give diverging answers (cf. section 4.4.4 above). To some extent, we observe reduced absolute differences. The major example is perhaps the reduction in absolute differences in infant mortality between occupational classes and between geographical regions. Thus, health inequalities are not unchanged. As an *overall* characterisation it seems nevertheless adequate to state that social variations in health, around the average level of health, have been relatively stable. The health advantages of the privileged positions have continued, during the transformation of societies from a situation of widespread poverty to a situation of relative prosperity.

The materialist/structural explanation can be interpreted as a prediction that if absolute material deprivation disappeared, population health in general would improve, and every major social position would have similar chances of obtaining good health. The first part of this prediction has in many ways been fulfilled, but the *second* part did not occur. A long series of trend reports have documented the continuing existence of health differences. These findings astonished many, as there was a widespread belief that generally better standards of living, welfare state arrangements, and comprehensive health services would lead to a substantial reduction of social inequalities in health.

Perhaps even more astonishing was the increased awareness of the so-called *challenge of the gradient* (Adler et al. 1994, Evans 1994: 5, Carroll et al. 1996: 27). The hypothesis that ill health was primarily produced by physical environments deviating negatively from some absolute levels necessary for health maintenance, would imply that among the upper strata, no important health differences should occur. However, large epidemiological surveys showed fine-graded differentials in most health indicators across the entire social hierarchy. The *Whitehall studies* have investigated large samples of employees in the British Civil Service during several decades. Employees are classified according to their precise location in the bureaucratic hierarchy. These studies have showed that not only were health differences present when

comparing the top administrative grade with the bottom clerical/office support staff levels, but also when comparing the top administrative grade with those at the immediate level below, the professional and executive level (Marmot 1996). Between these higher level occupational categories, with similar physical environments as regards their working conditions, relatively high incomes, very adequate material standards of living, and relatively high education, significant and systematic health inequalities in favour of the uppermost occupational grade were nevertheless found. Many other studies have indicated the same pattern. In most social hierarchies there seems to be a stepwise, gradual, deterioration in average health, from the top to the bottom, and this pattern appears also among categories which are placed far above the thresholds for material deprivation as they are usually conceived.

5.2.3. Reformulations of the basic model

Obviously, the basic model of the materialist/structural explanation of health inequalities has come under pressure by these findings. When studying the health disadvantages of particular sections of the populations – those who still are poor, also according to "old" definitions, or workers who still experience extremely harsh working conditions – these findings constitute no challenge. But as an *overall* explanation of society's health inequalities, the basic model appeared to be less credible.

Reformulations have therefore occurred. One alternative is to claim that the basic model is fundamentally right, but that more attention should be paid to the longterm accumulation of health hazards from the physical environments (Blane et al. 1998). Perhaps the working conditions and standards of living during the 1970s and 1980s did not vary between social positions in a way which could produce the current pattern of health inequalities, but the present inequalities do not necessarily reflect the circumstances of recent decades, but rather the life course exposures of the present adult population, accumulated over many decades – before affluence and welfare state developments.

Thus, the continuing health inequalities could be interpreted as a *cohort* phenomenon. They could reflect that large sections of the populations alive during the 1980s and 1990s had their health formed during the 1920s, 1930s and 1940s. At that time, variations in the physical environments were considerable, and the basic model of the materialist/structural explanation, with more emphasis on the life-time accumulation of exposures, could perhaps account for the continuing of health inequalities.

However, many studies show that health differences are present even among today's younger generations. Studies from the 1990s show hardly less

social variations in health among, say, people aged 25-35, than was found some decades earlier. The cohort explanation implies that the bulk of the inequalities would fade away, but hardly any evidence for diminishing health inequalities in the younger generation exists today.

Another alternative is to retain the assumption that the physical environments, in terms of physical working conditions and material standards of living, are the main health determinants, and variations in these circumstances are the main causes of health inequalities. The notion of threshold values is however abandoned. The basic model implied that some specific standards were sufficient for maintaining good health. When humans had physical environments deviating negatively from these standards, their health would suffer. When people had physical environments surpassing these threshold levels, no significant health benefits were believed to follow. *Enough is enough*: even better nutrition, even better housing conditions, even better clothing, etc., were not supposed to act positively on health.

Alternatively, it could be hypothesised that better physical environments will have a positive influence on health, whatever the previous level. It could be imagined that the association between the qualities of the physical environments and the ensuing health is *linear*. Therefore, also the present variations in physical environments continue to generate health inequalities.

The distinction between the notion of threshold values and the notion of a linear association between physical environments and health is seldom discussed explicitly. However, the continuing interest in the association between (relative) material deprivation and health, in spite of the rising average level of living, indicates that parts of the research field have an underlying notion that the association between physical environments and health is linear.

5.2.4. Towards a redefinition of the health determinants?

A *third* alternative is to reformulate the health determinants responsible for the health inequalities. The basic model emphasised the physical environments encountered by people in their work and in their standard of living. Health-related behaviours, such as smoking, physical exercise, etc. were generally not part of the model. Neither were psychosocial milieus given any large attention. A possibility is however to retain the structural determination part of the model, while at the same time to include other types of health determinants.

Thus, Blane et al. (1997), discussing the "materialist explanations of socio-economic mortality differentials", have argued that "Materialist explanations are

not confined to physical, chemical and biological hazards". The kernel of the materialist/structural explanation, it is argued, is not the material/substantial character of the health-producing factors, but rather the *social circumstances* producing the differential exposure: "materialist explanations refer to factors which are outside the individual's control and to which the social structure inevitably exposes some sections of the population or differentially exposes the whole population" (Blane et al. 1997: 385). Whereas the basic model puts relatively little emphasis on health behaviours and psychosocial circumstances, Blane et al. propose that such health determinants could as well be included in the materialist/structural explanation, because the essence of this explanation is not the particular type of health determinants, but rather the types of social processes which produce varying exposures to health-damaging factors.

This reformulation constitutes a bridge towards explanations which will be discussed in the following chapters. Before starting on them, it can noted that even Engels was not overlooking that behaviours could be responsible for the dreadful health conditions in the English working class. Engels acknowledged that many factors "enfeeble the health of a great number of workers, intemperance most of all". But according to Engels, drunkenness was an unmediated response to the harsh conditions of living. Structural victimisation was retained as the causal process. "Drunkenness has here ceased to be a vice, for which the vicious can be held responsible; it becomes a phenomenon, the necessary, inevitable effect of certain conditions upon an object of no volition in relation to those conditions" (1969: 133-134). Also as regards health-related behaviours, there is often no "volition" involved, but rather inevitable exposures, according to Engels. The explanation examined in the next chapter will usually, however, not accept this view of the social causation process.

6. The behavioural/lifestyles explanation

6.1. Overview and one example

6.1.1. The main characteristics

The themes of the behavioural/lifestyles explanation, and its differences to the materialist/structural explanation, can be understood as a reflection of changing historical and social circumstances. The materialist/structural explanation was rooted in 19th and early 20th century social conditions. The overwhelming facts of widespread material deprivation and premature deaths among infants, children and young people inspired its way of explaining health inequalities.

Although we will find many antecedents also for the behavioural/lifestyles approach, the post-war period constitutes a main background for this explanation. The health scene is the epidemiological transition. The advent of cardiovascular diseases, cancers, and chronic long-standing illnesses as major population health problems called for new ways of understanding how ill health was produced. Not absolute material deprivation, but the consumption patterns and the ways of living among the population in the post-war circumstances of growing prosperity came into focus.

Similar to the materialist/structural explanation, the behavioural/lifestyles explanation emphasises social causation processes, and main health determinants are found within the physical environments. However, both social causation and physical environments are conceptualised differently.

The explanation contends that people's lifestyles and behaviours emerge in consequence of the social positions they occupy. Each major social position has a typical profile of lifestyles and behaviours, distinguishing it from other social positions. A major theme is therefore how to understand the correspondence between location in the social structure and the behavioural/ lifestyles profiles. It is possible to explain this correspondence in terms of structural determination (cf. the last paragraph of the preceding chapter), but the main versions of this explanation formulate other social causation processes to account for this relationships. Usually, people are given a scope of action, a possibility to decide and choose, and a potential for varying adaptions. The underlying perspective is that people can amend their behaviours. Thus, compared to the materialist/structural explanation, individual and social network *agency* is more highlighted.

These lifestyles are connected to new notions of the central health determinants, which are found in other parts of the physical environments than those parts emphasised by the materialist/structural explanation. The *holy four* health behaviours (Lahelma et al. 1999) – smoking, drinking, eating, and physical exercise – are the main focus. The use of tobacco means that the smoker and his/her immediate milieus are exposed to health-damaging physical/chemical substances. Drinking alcohol in excessive ways can distort and destroy the health maintenance mechanisms of the body. Dietary choices and physical exercise are assumed to have effects on the body's susceptibility to disease and potential for good health. Often, also other types of behaviours are included, for instance accidents-proneness and the use of health services.

6.1.2. An example: Health behaviours and socioeconomic status

Aarø's thesis *Health behaviour and socioeconomic status* (1986) illustrates some typical aspects of this explanation. Its starting point is that in modern industrialised countries, disease and death are mainly caused by "lifestyle and environmental factors" (p. 1). Therefore, the task of public health and health promotion is to study how health behaviours and health-related lifestyles are distributed in the population, to analyse what predicts and causes health behaviours, and to develop effective interventions which can improve behaviours. The lifestyle approach is further assumed to be most fruitful for understanding *socioeconomic* differences in health. The aim of Aarø's study is therefore to examine the socioeconomic distribution of healthy and unhealthy behaviours in Norway, and to discuss why such distributions emerge.

In principle, health behaviour could include many types of conducts and actions, such as self-care, compliance to medical regimes, or social and political actions directed to health aspects of the broader environment. Aarø chooses to focus on a set of individual habits and behaviours which are assumed to affect the risk and probability of illness, accidents, disease, good health and well-being (1986: 11). The items reported in the survey include food habits, smoking, physical activity during leisure time, use of seat belts, sleeping habits, oral hygiene, and use of alcohol (p. 63). If such items were not correlated with each other, the hypothesis that health behaviours are important for explaining health inequalities would be difficult to sustain. The *dimensionality* of health behaviours is therefore an important question. Aarø finds that the items are moderately associated with each other. Factor analyses indicate three dimensions where items tend to cluster: *active health promotion* (food habits, seat belts, oral hygiene and leisure-time physical activity), *addiction and irregularity of daily life* (smoking, coffee drinking, habits as regards meals and sleeping), and finally *alcohol use* (p. 104-105).

The associations between socioeconomic status and these three dimensions are found to be statistically significant. Active health promotion, but also alcohol use, are more typical of higher status groups, while addiction/irregularity is more typical of lower status groups. Thus, Aarø's data show that the higher socioeconomic status, the more frequent are healthy behaviours. The exception is alcohol use, but moderate alcohol use is perhaps not an important contributor to ill health. Accordingly, although correlations are moderate, the study supports the health behaviour explanation of health inequalities. "Almost everything which may be labelled healthy is more common in high status groups" (p. 131). Given that these health behaviours are important health determinants, socioeconomic variations in health are partly accounted for, and the correspondence between socioeconomic status and health appears as *mediated* by health behaviours.

6.2. Social processes

6.2.1. More unhealthy behaviours in disadvantaged positions?

The explanation of social inequalities in health via the lifestyles and health behaviours of social positions would of course be implausible if social positions with more ill health did *not* have more behaviours deemed as unhealthy. Aarø's findings are in line with many other studies which document more unhealthy behaviours among lower occupational categories (for instance Blaxter 1990, Osler 1993, Uitenbroek et al. 1996, Lynch et al. 1997, Cavelaars 1998: 81-118). Similarly, the higher mortality among men compared to women has been linked to the health-damaging lifestyles characteristic of masculine ways of living (Annandale 1998: 139pp, Elstad 1999a: 28-29). Also, it has been supposed that the higher levels of ill health among divorced and separated people are a consequence of the relative frequency of poor health behaviours in these categories (Hemström 1996, Næss and Bergwitz 1991).

Thus, the empirical evidence indicates that those behaviours which are considered more unhealthy are more frequent in exactly those social positions which are disadvantaged as regards health. But *why* – what kind of social processes produce more unhealthy behaviours in these social positions? As asked by Lynch et al. (1997): "Why do poor people behave poorly?"

To understand this is a major issue for this explanation. From the standpoint of the individual, it will often be impossible to regulate the exposure to industrial air pollution, working conditions, or low income – those parts of the physical environments emphasised by the materialist/structural explanation. It is however different when tobacco smoking, an unbalanced diet, or lack of

exercise are the focus. Given the circumstances in present-day Western societies, it is hardly sensible to argue that ordinary workers are structurally determined to take less care of their teeth. A more sophisticated understanding than structural determination seems to be necessary for accounting for why people are differentially exposed to the health determinants considered to be most important by this explanation. In the following I will outline two important approaches to this question.

6.2.2. Health beliefs and the individualist approach

In this approach, the explanation for individuals' health behaviours is found in the ideas, conceptions and beliefs people have about illnesses and diseases, about the factors which enhance or deteriorate health, about the consequences of exercise, dietary choices and the use of tobacco or alcohol, about the appropriate way of handling symptoms and contacting the health services, and about the cost-benefit balance when health-related behaviours are chosen (cf. for instance Becker 1979).

Generally, this model locates the sources of health behaviours in the person and his/her knowledge, notions and beliefs. It has therefore been classified as an *individualist* and *cognitive* model (Barstad 1989: 19). There are many variants of this model. An example is Mechanic's model of the *help-seeking* process (Mechanic 1978: 268-269). Mechanic proposes that help-seeking depends on a series of factors, such as the visibility of symptoms, the extent to which symptoms are perceived as serious, whether they disrupt family life and work, available information, knowledge and interpretations of the symptom – and finally whether treatment resources are accessible in terms of physical proximity and monetary and psychological costs. Thus, actions taken by the individual are seen as the result of a series of cognitive factors: perceptions, motivation, preferences, information, knowledge. Corresponding models can be constructed as regards dietary choices, to smoke or not, etc. The models usually have an additional reference to external constraints: the unavailability of certain options may hinder people from choosing them.

A recurrent theme is the *health locus of control*, i.e., where people believe that the "power" over one's health is located. The underlying ideal is that people should believe that health is due to one's own actions. The opposite is to believe that health is created by external forces outside the control of individuals, for instance in people's inherent bodily constitution, in unavoidable environmental influences, or in a blind destiny. Often, the health beliefs approach focusses on whether people believe that their own choices

make a difference. If not, why should one select some types of behaviours because of health? The alleged *fatalism* of low status people, in contrast to the upper classes who more readily believe that health is influenced by own actions, is a typical issue (Aarø 1986: 135, Davison et al. 1992).

The production of health beliefs is furthermore examined. The cognitive characteristics of the model implies a focus on knowledge and education. The individual orientation of the model leads often to include personal traits in the model, for instance ability to absorb information, the degree of hostility towards others, and self-esteem.

6.2.3. Critical questions

The health beliefs model is perhaps the most common approach to the study of health behaviours. The model is in many ways designed by medicine's views on health and health promotion. Health behaviours are studied in ways typical of medical epidemiology (Aarø 1986: 3-5). Specific health behaviours are seen as functions of individual risk factors such as knowledge, information, beliefs, education, personal traits, gender, and occupational class. The health beliefs approach is closely linked to health promotion campaigns typical of our societies. Information and education are seen as the golden pathway to improve people's behaviours. Efforts are therefore made to redesign the curricula of schools, to advertisements, government-sponsored information to households, and the mobilisation of doctors as health educators and behaviour advisors.

Underlying the explanation is often the presupposition that people's ideas are wrong. "Beliefs" are identified with the unscientific, traditional, and prejudiced opinions of the lay public, deviating in various degrees from the scientific views of medicine (Bury 1997: 18-21). Studies of lay opinions, often a spin-off from the health beliefs approach, have however sometimes indicated a degree of coherence and "evidence-based" views which cannot immediately be characterised as irrational (Pill and Scott 1982, Bury 1997: 40-45). For instance, while the messages of the health promotion campaigns may be right when applied to the population level, doubts as to the necessity of mending one's own behaviours are not necessarily irrational. The strength of the association between behaviours and health outcomes at the individual level is rather small. The lay public knows that some do everything wrong and nevertheless live to the age of 90, others do everything right and die from heart attack before the age of 50. Medical epidemiology is aware of this: Rose calculated for instance that if 400 individuals used seat belts every day during

a period of 40 years, only one individual would avoid a deadly traffic accident, while no health gains ensued for the other 399 persons (Rose 1981).

Thus, the health beliefs model has been criticised from different corners. Its links to the medical establishment have been questioned, parallel to the mounting critique of medicalisation and the medical profession (Freidson 1970, Illich 1975). The current intellectual climate in the social sciences is not in favour of elevating one *belief* system (e.g., the medical one) as superior to other belief systems (e.g., lay opinions). Critics have furthermore questioned the model's way of understanding behaviour. The health belief model typically views "adult health behaviours (as) largely intra-individual phenomena which reflect some process involving free choice" (Lynch et al. 1997: 809). But are choices about health behaviours "free"? It has for instance often been discovered a discrepancy between the beliefs people have and their actual behaviour. Health beliefs explain only to a limited extent variations in behaviours (Williams 1995: 578). The *social patterning* of health behaviours constitutes another problem. Social positions are found to have particular profiles as regards health beliefs and health behaviours. This is suggested as the explanation of why social variations in health emerge. But social positions are collectivities of individuals, and can an individual-based model explain collective patterns? This criticism points towards other models for understanding health behaviours.

6.2.4. Lifestyles

An alternative to the health beliefs explanation is the *lifestyles* approach. It does not neglect that people make choices, but it views these choices as "situated within economical, historical, family, cultural and political contexts" (Lynch et al. 1997: 810). The freedom to choose is therefore in various ways formed by people's background, their resources, their experiences, and their group membership. Such processes form collective lifestyles, and the social and not only the individual roots of behaviours must be taken into account.

The lifestyle concept dates back to Weber and Simmel. For Weber, it implied a set of interrelated attitudes, habits, manners, behaviours, and consumption patterns, which "conveys a social meaning that displays ... the status and social identity" of the individual (Cockerham et al. 1997: 324). Thus, lifestyles are associated with status groups, and are principally a collective, rather than an individual phenomenon. Weber saw the emergence of lifestyles characteristic of a social group as the result of the interplay between life choices and life chances. Life chances represent resources, means, possibilities, as they are determined by the individual's position in the social structure. Life choices

represent the actual selection of behaviours and consumption patterns made by the individual, influenced by the constraints of his/her life chances.

Accordingly, health-related behaviours could be viewed as parts of more comprehensive lifestyles. They should not be seen out of context, but rather in their connections to the life chances pertaining to a social position, and to the ways of living and the typical consumption styles dominant in the status group. While the health belief model decontextualises behaviours and presupposes a tendentially unrestricted human agency, the collective lifestyle approach tries to disclose how structure and agency are interconnected. According to Williams, medical sociology has however paid little attention to how "the structure-agency problem in relation to health-related behaviour" should be understood (1995: 581). Two variants will be discussed below – the first one emphasises *structural constraints*, the second one emphasises lifestyles as *markers of group identity*.

6.2.5. Lifestyles as produced by structural constraints

Townsend, in a polemic against "the individualistic approach", argues that "Diet is profoundly influenced by cultural or local social customs, informal and formal education, the availability as well as price of goods in local markets, advertising, recipes and fashions recommended by the media, and decisions taken by farmers and the manufacturers of food products as well as by government" (1990: 383). "Similar considerations apply to other behavior, such as physical exercise, smoking, and drinking". Choice is therefore "shaped by powerful economic and social forces, the goods and facilities that are immediately available, and level of income". Low-income families, Townsend claims, "do not have incomes large enough to buy the kind of diet recommended for health".

Thus, Townsend presents choices as more or less determined by structural constraints. In his version, the distance to a structural determination formulation of how lifestyles emerge is not very far. The problem is not people's *willingness* to have healthy ways of living, but the actual range of options which are possible and appropriate. The constraints can be conceptualised in terms of material circumstances, as Townsend primarily does. They can also be seen in terms of customs and traditions. Deviations from prevailing norms can lead to social isolation, and will therefore commonly entail personal costs which are hard to bear. Generally, the behavioural options which are available and appropriate, are strongly influenced by the social context, which constitutes, in the words of Evans and Stoddart (1994: 44), a "powerful social conditioning".

The *culture of poverty* idea (Fitzpatrick and Scambler 1984: 58) is another variant of this approach, perhaps more suited to particularly marginalised groups. Their ways of living are seen as embedded in a tight network of customs adjusted to a poverty situation from which they cannot escape. Other variants of the structural constraints hypothesis would see behaviours as "natural" reactions to the circumstances. Exhausted manual workers cannot be expected to spend their leisure time doing physical exercise. The unemployed person smokes as a way of alleviating stress. A recent empirical study provides examples of the constraining effects of structures. It showed how changes in the working environment, giving more autonomy and decision latitude to the employees, led to lower levels of smoking (Landsbergis et al. 1998).

6.2.6. Lifestyles as markers of group identities

Rather than viewing health-related lifestyles primarily as reflections of constraints, they can also be seen as markers of group identities.

Williams (1995) treats lifestyles in this perspective, drawing on Bourdieu's theories of how social positions, i.e., locations in the social space characterised by their level of economic, cultural, and social capital, develop typical dispositions for taste, likes and dislikes, and social practices. This complex of attitudes and behaviours are rooted in the social structure. They are powerfully formed by the typical *habitus* members of different positions possess.

Habitus can be understood as "an acquired system of generative dispositions" (Bourdieu 1977: 95, Williams 1995: 585), or – more simplistically formulated – as internalised preferences, partly conscious, partly tacit and unreflected. They are produced by people's life histories and adjusted to their objective circumstances. The habitus characteristic of social positions acts to form typical lifestyles, and the health-related behaviours within these lifestyles should not primarily be considered as deliberate choices aiming at influencing health, but rather as parts of the overall "package" of lifestyles.

A recurrent question as regards Bourdieus' account of social practices in general, and of the typical lifestyles of social positions in particular, concerns how the opposition between structure and agency is understood, or – in other words – to what extent lifestyles are *products* of the circumstances or being *produced* by reflexive agents. Bourdieu's emphasis on practices as "taken-for-granted" points to the structuring power of objective circumstances. Various commentators classify Bourdieu among those theorists who above all highlight the determining aspects of social conditions: "structure is the dominant aspect of Bourdieu's concept of lifestyles" (Cockerham et al. 1997: 328).

According to Cockerham et al. (1997: 329-330), Giddens represents a position which presents lifestyles as conscious adoptions of ways of life. They aim at sustaining self-identity and are also ways of establishing membership in social groups. This opposition between Bourdieu and Giddens is not seen simply as disagreements, but rather as reflections of changing historical circumstances. While Bourdieu constructs his theories under the influence of mid-20th century conditions, Giddens reflects on the circumstances of "high modernity". Here, lifestyles are ways of expressing people's individuality, and lifestyles are more reflexively constructed.

The function of lifestyles as *distinguishing* ways of life, signalling how a social position is different from other positions, points to the struggle between positions for esteem, status and prestige in society. In this way, lifestyles can take the form of more or less conscious attempts to underline the dividing lines in society, and to make the social hierarchy more conspicuous. Lifestyles are therefore markers of social location and group identities, making the social divisions evident and reinforcing and reproducing the social hierarchies. Such lines of reasoning can for instance lead to the interpretation of the more healthy lifestyles in the privileged strata as not primarily aiming at health at all, but rather as ways of claiming superiority over other social positions.

Whether lifestyles are seen as products of structural constraints, or as rooted in the habitus of different class experiences, or as more actively produced ways of living, the common theme is always that lifestyles and their health-related behaviours correspond to the socioeconomic stratification of society. Various contemporary stratification theories deny exactly this correspondence. Waters (1994), for instance, claims that Western stratification systems are transforming historically. Economic class was the basis during 19th century industrialism, authority and organisational power were predominant during welfare state capitalism, while contemporary social divisions are rooted in adherence to "cultural items and processes, to idealised communities that focus on life-styles and value commitments" (Waters 1994: 296). Thus, lifestyles communities with loose or non-existing links to the economic social structures constitute main types of social positions, according to Waters. As regards the study of social inequalities in health, these views call for a focus on health differences between social positions belonging to a *lifestyle social structure*. I will not pursue this issue further. In passing it can be noted that even in societies with a plurality of non-economic divisions, such as Israel, lifestyles seem nevertheless quite markedly associated with the economic class structure (Katz-Gerro and Shavit 1998).

6.3. Health determinants

6.3.1. Antecedents

Similar to the basic model of the materialist/structural explanation, the behavioural/lifestyles explanation locates the main health determinants in the physical environments. Both nurture *biomedical* notions of how specific physical risks generate specific diseases. However, contrary to the former explanation's emphasis on work and material deprivation, the behavioural/lifestyles explanation views the main health determinants in terms of the physical environments emanating from people's unforced behaviours as regards diet, consumption patterns, customs, and leisure activities.

To view health risks from this angle has a long history. *Scurvy* can be taken as an example, a common disease of sailors, well known among the Vikings. During the middle ages, scurvy was explained by the habit of drinking too much sour milk on long sea voyages. The old Norwegian name of scurvy ("skyrbjúgr") is a combination of "skyr" (sour milk) and "bjúgr" (the typical swelling of the disease). Gradually, experience taught that citrus fruits could prevent scurvy. The founder of naval medicine, James Lind (1716-94), studied this systematically, and showed that scurvy could be prevented by relatively small changes in sailors' diet (Kock 1955: 194, 104).

Notions of the health consequences of diet, consumption patterns, hygiene, and stimulants, have been many through history – and often wrong, judged by today's knowledge. In Dickens' *David Copperfield*, one person claims that tobacco smoke cures his asthma. This view seems obsolete today, as does the idea that sleeping with a virgin would heal syphilis. With the growth of scientific *medicine* during the 19th century, a more sound knowledge basis was established, focussing on, among other issues, hygiene – how to avoid germs – and diet, in terms of the essential nutritional requirements of the body.

6.3.2. The focus in the second half of the 20th century

The view that physical environments created by typical health behaviours constitute a main health determinant came more into focus in the middle of the 20th century. This view arose, as said earlier, in connection with the epidemiological transition. Infections and infant and child mortality were declining. The major killers of middle-aged *men* – heart disease, cancer, accidents – became the main interest of public health (Kuh and Davey Smith 1993: 111). Whereas material deprivation was seen as the foremost cause of disease during the early 20th century, the main problems of the late 20th century

were said to be diseases of *affluence*. Not lack of food, housing, and income, but rather too much of the wrong things, available because of the prosperity enjoyed by large sections of the population in post-World-War II Western societies, was seen as the major problem.

Advances in medical epidemiology were a presupposition for the advent of this explanation. It focusses on the behavioural risk factors for the typical modern diseases. The health-damaging effects of smoking, lack of exercise, and diet were not immediately evident. They emerged after long-term exposures, often lasting for many decades. The health consequences of such behaviours could therefore only be established by large, longitudinal, epidemiological studies. The resources and technologies for making such studies were not available until well into mid-century, and the findings from such studies were the precondition for the advent of this explanation.

For instance, the reasons for the epidemic rise of lung cancer were debated for several decades. Some blamed the asphalt paving of roads which became common in the 1930s. During the 1950s, the role of tobacco was firmly established. A large study of cancer and smoking habits among English doctors at this time is considered one important breakthrough (Kuh and Davey Smith 1997: 25). Other large epidemiological studies advanced knowledge about ill health following from other behaviours. The role of physical exercise was for instance studied by the natural experiment provided by London bus transport. Bus drivers and bus conductors were similar in practically every respect – income, educational level, ways of living, etc. They differed however in the immediate physical circumstances of their work – seated, or running up and down the stairs on double-decker buses. Corresponding differences in heart disease were found (Blane et al. 1996b: 4-5).

Medical epidemiology examined possible health consequences of a series of other behaviours and consumption patterns. The study of diet was intensified, with a focus on fats and heart disease, but also with a great range of hypotheses about nutrition and various cancer types. The consequences of alcohol are more disputed. Drunkards had obviously higher risks, while moderate use is nowadays often regarded as indifferent, perhaps even healthy. The proper use of health services was also often seen to be essential for the maintenance of health, and illness behaviour and the determinants of physician utilisation were examined (Mechanic 1978: 268). The behavioural causes of accidents, in terms of individual risk-taking and personal characteristics leading to accident proneness were also studied (Holstein et al. 1988: 77). How to avoid infectious diseases – a long-term interest of this tradition – was constantly on the agenda,

and the *aids/hiv* epidemic starting in the 1980s highlighted in a new and dramatic fashion the relationships between behaviours and health.

At least since the 1970s, changes in health-related behaviours were widely accepted as the most important way of improving population health. The prevailing ideas can be illustrated by various estimations of the consequences of health behaviours on mortality. Based on the judgements of experts (i.e., selected physicians), it was proposed that 52 per cent of heart disease mortality, 37 per cent of cancer mortality, and 43 per cent of overall mortality, could be directly accounted for by health behaviours (Kringlen 1986: 177). Such estimations have guided public health efforts during the last decades. Typical in this respect is the Norwegian *Population Health Report* (Sosial- og helsedepartementet 1996). The section on main risk and causal factors for disease discusses only three topics: diet, alcohol/drugs, and tobacco. The implication was that the relatively high frequency of unhealthy behaviours among certain sections of the population was seen as a major cause of the existing health inequalities.

6.4. The Fundamental Cause Theory

Link and collaborators have advanced a theory of "fundamental causes", presented as a general understanding of why social inequalities in health emerge (Link and Phelan 1995, Link et al. 1998). It may be interpreted as a proposition that whatever historical circumstances, and whatever health determinants which are most important, social causation processes will emerge which generate health inequalities between socioeconomic positions. As I think there are many similarities between the "fundamental cause theory" and the behavioural/lifestyles explanation, I will discuss it in this chapter.

Link and collaborators distinguish between *proximate* and *distal* causal factors in explanations for health inequalities. Proximate causal factors are those factors believed to be the immediate causes of ill health: unhealthy behaviours, psychosocial stress, unhealthy working conditions, etc. Distal factors are the social positions themselves and their general social and material circumstances.

Link et al. criticise epidemiologists for their tendency to explain the higher frequency of disease among lower status groups by means of the proximate factors, i.e., immediate reasons for ill health. The typical procedure is to account for (or explain away) the health variations between social positions by means of the proximate factors. Thus, the relationship between social positions and the proximate causal factor appears as something accidental and external, as if there are no necessary links between the distal and proximate factors. The distribution

of proximate causal factors in the population should however be seen as produced by the distal factors, i.e., the social and economic circumstances of the social positions. The latter are therefore the *fundamental causes* of health inequalities. The distribution of proximate causal factors are produced by the distal factors, which are more basic in the causal chain.

Link et al. ask why the relationship between social class and health perpetuates, although the proximate factors responsible for health variations are continuously changing. No important modification of class differentials in health have occurred, in spite of changes in the health-producing circumstances: "... studies of the association between SES (socioeconomic status) and disease over the past several decades reveal an important fact – the risk factors mediating the association have changed" (Link and Phelan 1995: 86). The solution is to view the social class positions as a fundamental cause, while the distribution of proximate factors at any time should be interpreted as how this fundamental cause is materialising in the particular period.

What are the essential features of the distal factors? Link et al. choose to emphasise resources enabling people to meet, handle, and withstand the main risk factors which are acknowledged at the time. The important mechanism is "how general resources like knowledge, money, power, prestige, and social connections are transformed into the health-related resources that generate patterns of morbidity and mortality" (Link and Phelan 1995: 88).

This approach is illustrated by a study of cervical and breast cancer in the U.S. (Link et al. 1998). Mortality from cervical cancer is distributed in the familiar way, while breast cancer shows the opposite pattern: higher status women are more afflicted than lower status women. Why is this so? Link et al. suggest that screening influences the mortality of these diseases. The higher status, the more are women able to profit from screening instruments. As to cervical cancer, the *Pap test* has existed for many decades. Higher status women use it more, and mortality follows the familiar pattern. As to breast cancer, *mammography* has existed only for some years. Although higher status women use this more than lower status women, the health benefits have not materialised yet, and the opposite social distribution in breast cancer prevails. – The implication is that social inequalities in breast cancer deaths are expected to change in, say, a decade or two (the use of hypothetical future developments as part of the argument could be characterised as a weak point in this study).

The fundamental causes theory tries to formulate a generalised social causation explanation of health inequalities. However, it is most closely related to the behavioural/lifestyles explanation, as it seems primarily oriented

towards how the unequal distribution of resources in society enable or prohibit healthy behaviours generally, and as regards the health services in particular.

6.5. Criticism

6.5.1. Political questions: Blaming the victim?

The reason for the advance of the behavioural/lifestyles explanation of population health is sometimes claimed to be the political situation in many Western countries during the 1970s. Health services expanded in the post-war years and became more costly, but public finances were tight. Many doubted whether even more curative health services would be the solution to the dilemmas of population health (Hansen 1979). The slogan "Doing better and feeling worse" (Wildawsky 1977) was directed to the apparent paradox that although health generally and mortality rates in particular seemed to improve, the demand for health services increased. To redirect health policies towards prevention, stressing people's responsibility for their own health, seemed to be a way out of this dilemma. Preventative campaigns were more cheap than new hospitals. Ill health was claimed to arise from the ways of living of the population, rather than from the lack of public health services. Thus, the political pressure against governments was alleviated (Evans and Stoddart 1994: 37-45, Blane et al. 1996b: 4-7).

This has been characterised as a *blaming the victim* ideology. The individuals and their behaviours were seen as the causes of health problems. The sick persons had to bear not only their diseases, but also the shame of being themselves responsible for their unlucky fates (Crawford 1977, Townsend 1990). These moral overtones of the behavioural/lifestyles explanation have been the source of many critical attacks of the explanation, often suggesting that its scientific basis was fragile, and that its popularity in the ruling circles was primarily due to how it served the political needs of the establishment.

In particular the *health beliefs* model was the focus of many criticisms. Partly, this critique came from the social sciences, where many found shortcomings in the individualistic bias of this model. This criticism was spurred by the observation that many informational campaigns were shown to change the level of knowledge among the population, without resulting in significant and lasting behavioural changes (Williams 1995: 578, Blane et al. 1996b: 5, Syme 1996: 21-22). Although there are examples of successful changes of health behaviour (the Norwegian campaign for condom use in connection with the aids/hiv epidemics is one of them), these disappointing results raised doubts about the adequacy of individualist models for explaining behaviours.

6.5.2. Doubts about the health determinants

Also the views of health determinants were questioned. Are variations in behaviours such as smoking habits, physical exercise, dietary choices, and use of health services, the *main* reason for social differences in health?

The remarkable improvement in longevity in many Southern European countries and in Japan, generally faster than in the northern parts of Europe and America, has raised intriguing questions about the relationship between health behaviours and population health (Marmot and Davey Smith 1989, Marmot 1994, Bobak and Marmot 1996). In the south of Europe, the availability of high quality health services is hardly better than in the north, and negative lifestyles are hardly less widespread – so what are the reasons for the swiftly declining mortality rates in this part of Europe?

A particular question concerns the role of the health services. The sensible use of health services is also included among the behaviours supposed to create social variations in health. It is however a disputed issue how much the current available therapies influence population health (Fitzpatric 1997: 8-10). Most researchers view the general rise in standards of living as more important for improving population health than medical interventions (McKeown 1977, Hodne 1986). During the 20th century, life expectancy has increased by some 20-25 years in the West, but "generous estimates" find that the curative and preventative efforts of the health services at most can explain about a fourth of this improvement (Bunker et al. 1994, see also Wilkinson 1996: 30-31).

6.5.3. The epidemiological basis

This is connected to the present ability of medicine to understand *fully* the causes for ill health. Therapies can function without knowing the causes, as when psycho-pharmachological therapies alleviate psychic diseases, or chirurgic interventions cure some types of heart disease. Generally, however, the discovery of causes helps designing effective cures. It is often claimed that, as regards the common chronic and degenerative diseases – cancer, cardiovascular diseases, rheumatism, etc., – causes are only vaguely known. Although a long series of risk factors have been observed, with some, often slight, statistical association with particular diseases, it has been contended that "Yet few are the cases in which (the doctrine of specific etiology) has provided a complete account of the causation of disease" (Dubos 1995: 6). The same observation is made by Rose: "Most non-infectious diseases are still of large unknown cause" (1985: 34). The relatively adequate understanding of

the connection between smoking and lung cancer is perhaps a deviating example, and the causes of many other major diseases are less disclosed.

This is also suggested by studies which show that health behaviours (or risk factors assumed to reflect behaviours) cannot account for a large part of the *social variations* in the incidence rates of major diseases. Coronary heart disease is probably the most studied medical condition during the last 50 years (Syme 1996: 21). The British *Whitehall studies* (cf. section 5.2.2) found that the lowest category of employees in the Civil Service had a 2.7 higher age-adjusted risk of dying from heart disease during a 10-years period than the highest occupational category. Adjusting for behaviour-related risk factors – smoking, blood pressure, cholesterol, glucose tolerance – the relative risk was reduced by only about 25 per cent (Marmot 1994: 208, see also Wilkinson 1996: 65, Syme 1996: 25, Macintyre 1997: 738). Similarly, in the Norwegian *Oslo Study* 60 per cent of the variation in mortality from coronary heart disease, between the upper and the lowest occupational category, was still unexplained after taking the well-known risk factors (directly or vaguely related to behaviours) into account (Holme et al. 1980). Another Norwegian example is provided by Thürmer (1993: 79), who also demonstrated that behavioural risk factors at best gave only a partial explanation for occupational class differences in coronary heart disease.

Thus, *statistically* at least, social variations in heart disease could only to some degree be explained by behavioural factors. As to the variations in other major chronic diseases, the role of behavioural factors is even less clear. Some intervention studies brought the same message. The large and longitudinal Multiple Risk Factor Intervention Trial in the U.S. predicted the mortality improvements which were expected to follow from behavioural modifications. It turned out to be difficult to change behaviours, but "Even where behaviour was successfully changed, the ensuing improvement in mortality proved smaller than predicted" (Blane et al. 1996b: 6).

Accordingly, it seemed that behavioural differences between social positions explain *some* part of the social variations in health, but much less than supposed by the enthusiasts. From these doubts, different lessons could be drawn. Some advocated "the return of structure" (Blane et al. 1996b: 7), i.e., a renewed interest in the materialist/structural explanation. Others turned away from the preoccupation with the physical environments characteristic of both the materialist/structural and the behavioural/lifestyles explanations, and turned to psychosocial factors. This is the theme of the next chapter.

7. The psychosocial explanation

7.1. Main characteristics

Broadly speaking, first the materialist/structural and later on the behavioural/ lifestyles explanation were the main approaches in the research field up to the 1980s. The psychosocial explanation came more to the forefront during the 1990s. It also relies on social causation. We find various versions of how this social process is outlined. Some are quite similar to the structural determination characteristic of the materialist/structural explanation, others include human agency to a larger degree, but also new ways of understanding social causation processes are presented.

As regards the health determinants, we see two somewhat different tendencies. One sees health-related behaviours as a main health determinant. What makes this tendency different from the behavioural/lifestyles explanation, however, is that it understands the reasons for unhealthy behaviours otherwise, not primarily in terms of health beliefs or lifestyles, but rather as *reactions* to stress and strain and as ways of alleviating frustration. Thus, this variant is oriented towards understanding how social life generates frustrating experiences in the lower and disadvantaged social strata, and these negative experiences are the reasons for higher levels of short-sighted, self-destructive, desperate, and aggressive – and therefore unhealthy – behaviours.

In this chapter I will however mainly focus on the other tendency, which introduces a quite different understanding of the health determinants lying behind social inequalities in health. The materialist/structural and behavioural/lifestyles explanations conceive of health within the biomedical paradigm, emphasising the encounters between the human body and the physical environments. Contrary to this, the psychosocial explanation views health variations as products of social interaction. That humans are *social beings* is emphasised. Social environments and social relations are the central focus. Humans perceive, reflect on, and react to, their social environments. They communicate and interact within such environments, and the resultant *psychological* reactions are a main reason for health and ill health.

Thus, in a way the psychosocial explanation reorganises the dividing line between the social sciences and medicine. The study of how health is created becomes more integrated with social science. The concepts of sociology, anthropology, and psychology become fundamental for explaining health.

Within the materialist/structural and behavioural/lifestyles explanations, the role of biomedicine was to discover how physical substances acted on the body. Within the psychosocial explanation, the role of biomedicine is more often to understand the medical consequences of *psychological* phenomena, and how bodies' physiological, neurological, and immunological processes are influenced by psychological processes.

This notion of how health is created is a distinguishing characteristics of the psychosocial explanation. I will start by discussing this health determinant, before I turn to the social processes put forward by this explanation.

7.2. Health determinants

7.2.1. Background: the context of late 20th century

The psychosocial explanation of health does not deny the potential physical environments have for directly influencing health. Poor nutrition, poor housing, and unhealthy working conditions are accepted as health risks. Behaviours which create dangerous physical environments (filling the air with tobacco smoke, for instance, or harming own and others' health by reckless driving or violence) are furthermore readily accepted as relevant. So far, this explanation does not reject well-established biomedical knowledge. But, as Renaud writes, the "struggle against bad life-styles (is) necessary yet insufficient". Even if the whole population had healthy physical environments, nevertheless "the health gradient between the various social classes might well remain" (1994: 320, 322), because there are other causal chains influencing health than those which consist of the human body – physical environments interaction.

Thus, it is suggested that the health determinants which are highlighted by the former explanations, no longer have the same relevance. The psychosocial explanation claims that historical change has transformed the determinants of health variations. In the affluent countries, the potential for improving the average level of health through alterations of the average level of physical environments *per se* has largely been exhausted. Generally, it is supposed that people have sufficient food, acceptable housing conditions, and seldom directly harming physical conditions at work. Under such circumstances the social environments and psychosocial influences become essential. As Freund writes (1982: 6, referring to Antonovsky 1979): "... when a certain standard of living is reached psychosocial influences replace biochemical and physical stressors as the main determinants of health". Similarly, Wilkinson (1996) supposes that in present-day rich countries, no general improvement of life expectancy can be expected to follow from even higher average material prosperity.

7.2.2. The focus on information and mental reactions

What are these psychosocial influences, and how can they influence health? Generally, the hypothesis is that people's *mental reactions* to their life experiences and to their social surroundings have important consequences for their health. Thus, information is a crucial aspect. People receive information from outside which is processed by the central nervous system, and from this, processes may develop which influence health. Usually, this is easily accepted as regards mental health (although there are of course biomedical explanations for mental health as well). Despair, hostility, rage, disappointments, etc., are feelings incited by experiences, and it is generally accepted that depression, anxiety, and psychotic conditions may follow. The psychosocial explanation goes however further and argues that also *somatic* disease is not only possible, but even frequent, consequences of such mental reactions.

Lay opinion has few difficulties in accepting a link between mental reactions and somatic disorders. In folk wisdom, the loss of one's beloved, or the frustrations experienced when one's life ambitions are crushed, are often regarded as causes of physical ailments. Such links can be understood on the basis of familiar experiences that information from outside may directly provoke bodily changes: one blushes when caught in shameful acts, thrilling experiences are accompanied by uncontrollable heart beats, and one pukes when confronted with disgusting events.

However, in medical science it is a disputed question whether and to what degree somatic diseases can develop in this way. What are the *biological pathways* (cf. Kelly et al. 1997) that make the psychosocial link plausible and even important for the emergence of somatic disease? Conventional biomedicine has its doubts. Nevertheless, it is seriously studied by rapidly expanding medical fields such as psychoneuroendocrinology and psychoneuroimmunology. Specialised and complicated topics are involved, beyond what can reasonably be discussed here. What I believe is a widely accepted and perhaps fruitful model will be outlined below.

7.2.3. Mental arousal and bodily deregulation

The living human body can be seen as several dynamic *body systems* – the nervous system (transmitting signals around), the cardiovascular system (regulating and pumping blood), the digestive system (transforming food into energy), the endocrinological system (involved in the secretion of hormones), the immunological system (not only killing invading microbes, but generally directing "the repair of damaged tissue", Evans et al. 1994b: 169), and so on.

The living body and its healthy state depends on the balanced functioning of these various systems, and therefore on the mechanisms which regulate them.

Most of these systems function autonomously, in the sense that it is usually not up to our volition how they operate. We may of course decide to think about something, or decide to move our limbs, but the heart beat rate, the blood pressure, the production of hormones, the digestive process, the functioning of the immunological apparatus, etc., go on, without any conscious decisions. The psychosocial explanation argues that these systems of the organism are closely interrelated, and moreover influenced by the nervous system. Thus, when something goes on in the mind, it may have consequences, without our conscious knowledge, for the functioning of other bodily systems. Evans et al. (1994b: 168) explain: "Perceptions of the external world, detected and interpreted by (the nervous) system, lead to electrical and chemical responses that in turn trigger responses in other systems".

The ability of mental responses to modify, change, alter – and even *disturb* – other bodily systems is a crucial part of the psychosocial explanation of somatic disease. The origin of these ideas within the scientific community can be found in the studies of Walter B. Cannon during the 1930s, later pursued by his pupil Hans Selye (Freund 1982: 41). Selye studied how external *stressors* were followed by endocrinological changes, for instance the secretion of various hormones. External stressors consist of any kind of situation which is perceived as challenging or demanding: danger, a university test, the duties and tasks involved in one's job, unemployment, the loss of a relative. External stressors require *adaption*. The organism has to confront the external stressor, by efforts to adapt, to cope with the situation, and to find adequate responses. This means an arousal of mental functions and involves a mobilisation of the organism, often reflected in alterations of bodily systems: muscles get tense, the heart beat rate accelerates, the production of adrenalin changes, etc. This mobilisation of the body when confronted by external demands is called *stress*, or the stress response (terminology is sometimes vague, but usually, stressors are the external situation while stress is the responses of the mind and body).

Generally, this mobilisation of the body and its systems is not unhealthy. Rather, it is necessary for the ability to function adequately under various circumstances. Normally, these changes of bodily systems in response of external stressors last for a while, until successful adaption has occurred. However, the healthy functioning of the body requires that the stress reactions are *turned off*. The mobilised body involves unusual and perhaps extreme endocrinological and immunological states which, if continuing, may harm the body. "The ability to turn off the stress response may be almost as important

for the long-run survival of the organism as the ability to turn it on for short-run survival, because if the (stress) syndrome is not properly turned off, and if the level of circulating cortisol remains inappropriately high, the result of prolonged failure to carry out normal maintenance and repair is accumulating damage to a number of different types of cells" (Evans et al. 1994b: 173).

Accordingly, the mobilisation of stress responses, when *not* followed by the proper restoration of the normal functioning of the endocrinological system, the immunological system, etc., can be harmful. It may entail dysfunctional states which can provoke or facilitate the development of somatic disorders. Health maintenance requires that the body is allowed to restore the balanced functioning of its various systems. "Unregulated arousal", on the other hand, can lead to states of "bodily deregulation" (Freund 1982: 43, 39), with possible disease consequences of many kinds: lasting elevated blood pressure, digestive problems, and poorly working immunological defences.

Approximately in this way, the psychosocial explanation supposes how somatic disease may follow from external stressors. Mental arousal upsets the *homeostasis* of the body, i.e., the balance of bodily processes. If external stressors produce mental arousal too often or too strong, or if it happens constantly and more or less uninterrupted, the capacity for restoring normal bodily processes is impaired. The stressors are exhausting and exceeding the ability to adjust, and the normalisation process is not functioning. Therefore, lasting bodily deregulation may occur, with a potential for somatic disorders.

This explanation of somatic disease, from external stressors to processes of the mind, followed by deregulated processes in the body, also raise the question about how mental processes arise. Some suggest that the reasons for mental arousal are universal, i.e., that they are deeply rooted "in a more biological, universalistic, frame of reference" (Williams 1998: 122). To view mental arousal as *socially constructed* is perhaps more widespread. Mental reactions are influenced by norms, preferences, and prevailing types of evaluations and classifications of circumstances and events. Whether or not a certain experience or external event produces drastic or trivial mental arousal depends on the society's and the culture's normative systems, i.e., on common ideas about what constitutes defeat or success, and about what should incite feelings of disaster or indifference. Whereas external demands may be objectively described, mental arousal is a subjective state, not simply reflecting the objectivity of the stressor, but conditioned by many personal, social and cultural factors. Thus, the *social conditioning* of mental arousal becomes an important problematic within this explanation (Williams 1998).

7.3. Social processes

Thus, the psychosocial explanation views social inequalities in health as a result of how the disadvantaged social positions commonly experience higher levels of unhealthy stress responses. Such unhealthy stress is usually seen as having diffuse and unspecific negative consequences and may therefore lead to many types of health problems: poor mental well-being, bodily symptoms, and many types of somatic disease. The *general susceptibility* view is therefore closely related to the psychosocial explanation. In this section, I will discuss attempts to understand why elevated levels of stress occurs in the disadvantaged social positions. Why do unprivileged positions more often experience stress?

Various hypotheses about this have been presented. They are often overlapping and complementary. A distinction can be made between approaches which primarily focus on individuals and their locations in the social structure, and approaches which emphasise the societal level. I start with two individual models (the *vulnerability* hypothesis, and the *epidemiology of social stress* model), continue with a section on *social hierarchies, control and relative deprivation*, and finally discuss the societal approach: the *social cohesion and social capital* model.

7.3.1. The vulnerability hypothesis

According to the vulnerability hypothesis (Aneshensel 1992: 23-27), the basic process is not so much that the disadvantaged social positions encounter more external hardships and life difficulties (as in the epidemiology of social stress model, discussed in the next section). Rather, health differences due to stress occur because the disadvantaged social positions lack the resources necessary to handle and adjust to external stressors. They have less *coping resources*. They are more poorly equipped with the assets which may be activated when confronted with difficulties. They are more vulnerable when something happens, because they lack "generalised resistance resources" (Lundberg 1997: 823). The focus is therefore on differences in *stress reactivity*, or, as Aneshensel technically puts it: the major mechanism is "group differences in the coefficient for psychological distress regressed upon a stressor" (1992: 23).

Various processes can account for such vulnerability differentials. *Attachment theory* focusses on the early social environment, primarily on the "infant's emotional experience in the relationship with the caregiver" (Fonagy 1996: 131, also Killén 1999). If love, care, and closeness are deficient, the infant develops distorted patterns of attachment. Distrust, anxiety and

diminished feelings of security may be the result. Such patterns of attachment may endure for a long time, and people carrying with them the insecurity from the early years react more strongly to threats and challenges encountered later in life. They become less able to cope with external stressors and will regularly experience higher levels of unhealthy stress.

Due to the more difficult social situations in lower status families (less income, less resources in terms of education, etc.), such families are often less capable of fostering sound patterns of attachment in their children. Antonovsky's *sense of coherence* (Antonovsky 1979, 1987, Lundberg 1997) is in many ways parallel to attachment theory. Antonovsky argues that a personal orientation which views life as meaningful, comprehensible, and manageable is a salutogenic factor. It helps the person to avoid negative stress and to stay healthy. The sense of coherence is created and influenced by experiences and the availability of resources during the entire life course.

Differential vulnerability between social positions may therefore emerge because of differences in acquired personality structures. They may be formed by the social circumstances in early life, but may also develop during one's entire lifetime. Vulnerability may also depend on other resources, present or absent in the situation of the individual, which enable him or her to resist stressors and to adapt satisfyingly to external demands. A form of such resources is *social support.* Social networks – friends, spouses, close relatives, work mates – constitute supportive social relationships, improving people's ability to withstand negative stress: they give emotional support and practical help (Pearlin 1989, Aneshensel 1992: 25). Differential vulnerability may also depend on even more tangible resources. Education, income, and material resources of every kind may act as buffers against stress, thereby influencing vulnerability, and such resources are of course distributed among social positions corresponding to the way health is distributed.

7.3.2. The epidemiology of social stress

Rather than focussing on differences in vulnerability, the epidemiology of social stress approach argues that what primarily differs between social positions is the amount of external stressors they encounter. External stressors are conditions demanding coping, adjustment and adaption. This approach claims that disadvantaged social positions are more frequently confronted with stress-producing circumstances, and this results in higher levels of ill health.

Such external stressors are of very varied types. They are often divided into *life events* and *chronic stressors* (Pearlin 1989, Turner et al. 1995). Common

types of life events are death of a close relative or friend, the experience of theft or "trouble with the law", an unwanted pregnancy, divorce, a fire in one's house, sudden unemployment or financial crisis, or failing a school or university test. Such events often appear suddenly and unexpectedly, and require a rapid reorientation and adaption. Chronic stressors are of a more permanent, longstanding type, constantly influencing the state of the mind. Simply because the stressor cannot be avoided or stopped – no successful relief from the stressor can be found – it constitutes a longterm, chronic stressor: perpetuating economic problems, unsatisfactory working conditions, too much work pressure for a long time, longterm harassment from a boss you cannot escape from, and lasting conflicts with your spouse or worries about your children.

The link to how social inequalities in health emerge is found through the study of the epidemiology of social stress. "Many stressful experiences ... don't spring out of a vacuum but typically can be traced back to surrounding social structures and people's locations within them. The most encompassing of these structures are the various systems of stratification that cut across societies, such as those based on social and economic class, race and ethnicity, gender, and age" (Pearlin 1989: 242). Various life events – for instance the death of a close relative, unemployment episodes, or sudden economic difficulties – are more frequent among lower strata, but the social differences are even more marked as regards chronic stress (Turner et al. 1995). Aneshensel (1992) distinguishes however between *random* and *systemic* stressors: the former occur more or less randomly, across social structures, while the latter are significantly associated with the locations in the social structures and occur more frequently in those social positions which commonly have more ill health.

7.3.3. Social hierarchies, control and relative deprivation

Both the vulnerability model and the epidemiology of social stress model are compatible with a view that stress diminishes the higher one is located in the social hierarchy, and increases the more one approaches the lower end of the social scale. However, these two models are not reflecting that different social positions are not only characterised by what pertains to the position in itself, but also by the position's relations to other positions. Thus, that people in a particular position are exposed to some particular level of stress is highlighted, but the corresponding fact that other positions are experiencing less or more stress-producing circumstances are not reflected.

Other researchers advocate psychosocial explanations which emphasise that social positions are not isolated categories, but parts of *hierarchical* systems. Thus, the determinants of health of a particular position are not only

found in the circumstances of that position, but also in its relations to other positions and in its relative location in society. Generally, the consequences of ordering people in social hierarchies are considered. One impetus for this has been the "challenge of the gradient" (cf. section 5.2.2): the observation that average health corresponds to the level in social hierarchies, across the entire hierarchy, and not only as differences between top and bottom. Therefore, Evans (1994: 6) suggests that there must be "something that powerfully influences health and that is correlated with hierarchy per se". This "something" is seen by Evans as psychosocial, but what is it, more specifically?

For Syme (1996, 1998), the crucial concept is *control of one's destiny*. The critical variable is the extent to which one can influence the events that affect one's life. The ability to do this is associated with problem-solving skills and access to resources (Syme 1998: 498), and these factors are, by and large, dependent on the hierarchical position. The higher located, the more skills and resources are available, and there are less people even higher in the hierarchy whose power constrains one's own control. Syme argues that the concept of control is a common denominator of various other terms often used in psychosocial explanations: mastery, self-efficacy, locus of control, powerlessness, controllability, and learned helplessness (1996: 28). Syme substantiates this hypothesis by various studies, for instance by Karasek and Theorell's (1990) findings about the health-deteriorating effects of high job demands and little decision latitude, i.e., little control in the work situation.

Thus, the scope of control perceived and realised by the individual, and the extent to which hierarchical structures and distributions of resources allow or constrain control, is the major mechanism, according to Syme.

Somewhat different is Wilkinson's emphasis on *relative deprivation* (1996, 1997, 1999). Here, the point is that "socioeconomic differences in health within countries result primarily from differences in people's position in the socioeconomic hierarchy *relative to others*" (1997: 593, my emphasis). Not lack of control in particular, but rather the awareness of people that they are deprived, not in absolute terms, but relative to people higher in society, is highlighted by Wilkinson.

This relative deprivation can affect multiple dimensions: power, status, income, material standards, good-quality houses, prestige, and qualifications. The ranking in society of people according to such dimensions regularly involve *social comparisons*. People compare their lot with others, in particular with those higher up, and this social comparison may incite feelings of frustration, inadequacy, inferiority, anxiety, i.e., chronic stress. Social comparisons may incite hopelessness and aggression, leading to self-destructive behaviours and

violence, and an important aspect of relative deprivation is that it conveys the message that those deprived are "less successful, less attractive, less intelligent, and socially inferior" (1999: 533). Thus, the problem with social hierarchies is that it "presents itself as if it were a hierarchy of human adequacy", and in general, hierarchies entail corresponding distributions of social stress and result in corresponding distributions of stress-induced illness and disease.

7.3.4. Social capital and social cohesion

The emphasis on social hierarchies shifts the focus from the situation of each social position to how social positions are parts of social structures. A further development towards the societal level is represented by hypotheses about the consequences of social capital and social cohesion.

The starting point for this development was studies which indicated that the average level of health in a country was associated not only with the general socioeconomic level in society, but also with the degree of material inequality prevailing in that society. Others have reported such findings (Rodgers 1979, Waldmann 1992, Wennemo 1993), but in particular Wilkinson has discovered such patterns and linked them to explanations of social inequalities in health (1993, 1994, 1996). Together with collaborators he has tried to develop a comprehensive theory of health inequalities between and within societies in line with these findings (cf. Wilkinson 1996, 1999, Wilkinson et al. 1998, Kaplan et al. 1996, Kawachi and Kennedy 1997, Kawachi et al. 1997).

The presupposition of this explanation is that absolute deprivation has ceased to be a major cause of ill health and premature mortality in the affluent countries. Sufficient material standards for health maintenance have been attained by the majority. However, this does not imply that inequalities in society have disappeared. Marked inequalities as regards levels of living and income are continuing. The amount of inequality does however vary between communities, regions, and countries. Some are characterised by large inequalities, while other societies are more equal in the sense that material differences between the social positions are not so marked and significant.

What are the consequences of these material (and in particular, *income*) inequalities? Partly, the consequences are those described in the preceding section. Social hierarchies, and the ranking of people and social positions according to their level of income and status foster relative deprivation. The sharper the social divisions, the more health-damaging consequences will be experienced by those located at lower levels in the social hierarchies.

But apart from this, income inequality has also consequences as regards the *modes of social relations* prevalent in society: "income inequality is

strongly and systematically related to the character of social relations and the nature of the social environment in a society" (Wilkinson 1999: 526). The larger income inequalities, the more will social life suffer from disintegration and lack of social cohesion. When the inhabitants are more equally ranked, this creates a basis for co-operation, trust, participation in civic affairs, and mutual respect. Inequality, on the other hand, fosters lower levels of trust and higher levels of hostility, widespread frustration, and feelings of inferiority and lack of dignity. Thus, the more inequality, the "whole distribution of social relations is shifted towards the more aggressive ... and conflictual" (Wilkinson 1999: 527).

The level of *social cohesion* is thus one side of the coin, strongly affected by the amount of social inequality in income, status and authority. The level of *social capital* in another, related, aspect. Social capital refers usually to the networks, norms and trust that prevail in society, which enable or hinder members of society to act together in order to pursue shared objectives (Wilkinson 1996: 221). Kawachi et al. (1997) define social capital in terms of the level of civic engagement (participation in community activities, elections, and voluntary organisations) and the level of trust in social institutions and in other members of society.

Thus, the hypothesis is that (large) social inequalities generally, and income inequality in particular, undermine the levels of social cohesion and social capital in society. Inequality leads to circumstances where people live in a social environment characterised by distrust, hostility, aggression, frustration, and disrespect, with less sense of belonging to the wider society, and less community feelings. Such social environments create, in themselves, chronic stress which may harm the health of everyone in society, no matter where in the hierarchies one is located. But income inequalities which induce low levels of social cohesion and social capital afflict in particular those who are relatively underprivileged: "social breakdown and conflict ... tends to be concentrated in the most deprived areas" (Wilkinson 1999: 529). Here, chronic stress will be even more widespread, as a synergic effect of both low levels of cohesion and high levels of frustration because of relative deprivation. Moreover, health damages from violence and homicide will be particularly evident. Typically, the focus on social cohesion and social capital has led to a special interest in health inequalities emerging from violence, aggressive behaviour, and crime (Wilkinson et al. 1998, Wilkinson 1999: 529-531).

The social cohesion and social capital approach develops the psychosocial explanation for social inequalities in health in a distinct direction. It introduces *social facts*, in a Durkheimian way (cf. Muntaner and Lynch 1999), as an

explanation of health inequalities. Moreover, it integrates hypotheses of how health inequalities *within* societies are created with hypotheses of how differences in health levels *between* societies are produced. It points to a solution to the apparent paradox that societies with lower material standards may nevertheless have better average health than societies with higher material standards. The reason is the mediating factor constituted by the magnitude of social inequality in itself, and its resultant levels of relative deprivation, social cohesion and social capital. Lower average levels of income may be compensated by higher equality, and population health may improve in comparison with societies high both on average income but also on inequality.

7.4. Debates

7.4.1. Are psychosocial influences a main health determinant?

The advent of psychosocial explanations for the general level of health in modern societies and for the continuing social patterning of health has been claimed to be a "paradigm shift" (Evans et al. 1994a: ix). Studies of the relationships between stress and health date back many decades, but recent years have seen an intensified interest in psychosocial determinants. The idea that health in contemporary societies is grossly influenced by social relations and psychosocial processes has become widely accepted. Williams, for instance, contends that is a *fact* "that once a certain (material) standard of living and 'epidemiological' thresholds has been reached, other, more diffuse and intangible, factors take over as the major determinants of 'socially patterned' disease in advanced Western societies" (Williams 1998: 121).

In spite of such declarations, the association between psychosocial factors and disease is a disputed topic. On the one hand, a large number of studies indicate that there is such an association. On the other hand, the associations which are found are often weak and sometimes difficult to replicate. As to the association between life events and disease, Aneshensel notes for instance that "the deleterious health effects of life changes are of consistently modest magnitude" (1992: 17). When indicators of vulnerability and social stressors are included in medical epidemiology studies, they often contribute significantly, together with behavioural or material factors, but their contribution is seldom large.

Thus, a solution to the "mysteries" of ill health and disease is not immediately found by introducing psychosocial indicators. Methodological problems – how to measure, how to define health (as specific disease entities or as more broad, general health measures?), how to separate causes from confounding factors (Davey Smith and Phillips 1992) – are numerous, as are practical and certainly ethical problems. The interrelatedness between material,

behavioural and psychosocial factors complicates the separation of psychosocial influences from other factors. Some suggest that psychosocial effects are overrated, and that health should rather be understood as a consequence of longterm influences from material factors (including those with a behavioural origin) interacting with psychological factors (e.g. Davey Smith 1996).

Uncertainty about many of the hypothesised associations between biological processes and vulnerability, social stressors, and hierarchical positions often leads proponents of this explanation to search for evidence in animal studies. The association between hierarchical positions and endocrinological functions among baboons is a favourite example (Evans et al. 1994b: 171-175, Wilkinson 1996: chapter 10) – it should however be noted that also human studies have found similar results (Brunner 1997).

A particular problem encountered by investigations of the role of psychosocial factors is how to understand mental arousal. A distressed mind is supposed to be the factor preceding harmful distortions of the endocrinological and immunological system, but how does the distressed mind emerge? *Emotions* can be conceived of as a link (Williams 1998, see also references in Elstad 1998b). The wider social structure generates objective stressors, but only when stressors mean something in emotional terms, can they be expected to lead to mental distress. In the further development of the psychosocial explanation, the study of emotions will perhaps be central.

7.4.2. Debates about income inequality, social cohesion, and health

The hypothesis developed by Wilkinson and associates (section 7.3.4) supposes, basically, a causal chain from income inequality to social cohesion and social capital, and from these factors to levels of health and health inequalities. Different units of analysis have been considered. Wilkinson focussed on the health differences between countries (Wilkinson 1994, 1996: chapter 5). The U.S., with its large population divided between 50 states, is of course a suitable testing ground for the hypothesis. A number of ecological studies of American states have been made, most of them supporting the hypothesis (Kaplan et al. 1996, Kennedy et al. 1996, Kawachi and Kennedy 1997, Kawachi et al. 1997, Wilkinson et al. 1998, Kahn et al. 1998). Also in Britain, using electoral wards as the geographical units, patterns are found which are compatible with the hypothesis (Boyle et al. 1999).

Thus, an association between income inequality and average population health, after adjusting for the average income level, has been demonstrated for various types of geographical units: countries, U.S. states, and smaller areas. This gives credibility to the hypothesis. An unanswered question is whether

this association can be found, regardless of how small the units are. Soobader and Leclere (1999), using U.S. data, found less evidence for the association when the smallest geographical units were analysed than when data were aggregated to higher geographical levels. This draws attention to questions about how the relevant social processes are understood. If the mediating factor is social cohesion, *where* is social cohesion generated – in the neighbourhood, the local community, cities, regions, or at country level?

The support for the hypothesis is nevertheless not unanimous. Judge, for instance, claims that the results are dependent on relatively arbitrary choices as to how income inequality is measured (Judge 1996, Judge et al. 1998). Gravelle (1998) argues that Wilkinson's results may simply be a statistical artefact. The effect of income on health is probably *curvilinear*, so that increasing income in the highest ranks of the income hierarchy gives diminishing health benefits. For societies with the same average income, societies with larger income inequality will have both more very poor people and more very rich people. As the adverse health effects of reduced income in the lower ends of the income hierarchy are stronger than the beneficial health effects of increased income in the higher income levels, it follows automatically that average population health will be lower where income inequality in higher. Similarly, Ecob and Davey Smith (1999) argue that the fundamental relationship is between income and health at the individual level, and that this relationship "is compatible with worse health in countries with greater income inequality, without the need to postulate any direct effect of income inequality itself". When income differentials are large, large sections of the population will have particularly small incomes, and this may directly account for the lower average level of health.

Other discussions address the social processes which are involved. Why does income inequality generate social cohesion, and is high social cohesion always a health-promoting factor? The hypothesis that income equality fosters social cohesion may be too simplistic (Elstad 1998b: 611). Generally, to what extent cultural and social life mirrors economic structures is a grand sociological theme, and it may be said that Wilkinson's account of this is sketchy. Lynch and Kaplan (1997) claim that some very cohesive societies have been rather unhealthy. Some indicators of social cohesion may in fact reflect high levels of social control, and are forced participation and strict norms compatible with beneficial psychosocial circumstances? According to Muntaner and Lynch (1999: 61), whether social cohesion mediates between income inequality and health is "mostly an untested hypothesis". – The further development of the inequality – social cohesion – health hypothesis will supposedly depend on its ability to answer such questions convincingly.

8. Social mobility explanations

8.1. Overview, examples

8.1.1. Introduction

Chapters 5 – 7 have examined variants of social causation explanations for social inequalities in health. Now I will turn to the "opposite" social process – social mobility – and its possible relevance for explaining health inequalities. The role of social mobility in generating social inequalities in health is a long-standing issue. The 19th century sociologist Eilert Sundt, for instance, supposed that illness could be "an important cause of impoverishment" (see the Introduction, p.1), and suggested that health problems not seldom initiated a process of social degradation, leading to a higher proportion of persons with ill health in the lower social strata.

Another early sociologist who studied the relationship between health and social mobility was the Russian-American *Pitirim Sorokin* (1889-1968). Using historical data, Sorokin noted that those who rose from humble origins to higher social strata regularly had good health, a high stature and a strong physique. He showed that popes, presidents and millionaires (!) generally lived much longer than the average population (Sorokin 1964[1927]: 259), and suggested that this was not only because of better environments, but also because of their healthy constitution. According to Sorokin, "Climbers, as a rule, must be strong people. Without endurance, energy, force, and health, they cannot achieve, and thereby become prominent and promoted" (1964: 269). Their superior health, if not actually causing upward mobility, was at least a presupposition for it. They had to live long, for it took decades to become included in the top circles. Physical strength was required for the strenuous climbing to the top, and upward mobility would be effectively obstructed by ill health and of course by premature death.

Thus, the social mobility explanation has always existed alongside, and constituted a *challenge* to, social causation explanations. It has often been regarded as a contributory understanding of why social inequalities in health emerge, but also sometimes as the main explanation. Mostly, it has nevertheless played a secondary role within the research field. However, as social mobility is a characteristic feature of society, the research field has had to consider to what extent social mobility is involved in the generation of

social inequalities in health. This is necessary not least because of the frequency of social mobility in contemporary societies.

As to the marital structure, for example, numerous transitions from being unmarried to being married occur, not seldom followed by divorce (and remarriage). During the 1990s in Norway, more than 20 per cent of marriages were expected to last less than 10 years (Statistics Norway 1998a: table 80). Geographical mobility is also frequent. Data from the Norwegian Health Survey 1995 show that, when dividing Norway into six regions, some 20 per cent of the adult population (aged 25-74) currently lived in another region than the one they resided in during childhood. Likewise, mobility between social positions in the occupational structure happens often, in particular as regards intergenerational mobility. A common finding in Norway and similar countries is that only 30-40 per cent of adult men belong to the same main occupational category as did their fathers (Ramsøy 1977: 109, Li and Singelmann 1998: 323, see also Ringdal 1994, Eriksson and Goldthorpe 1992, Elstad 2001).

Thus, social positions in most social structures will often consist of many who have been recruited from other positions. If mobility is *health-related*, so that stayers and movers vary as to their health characteristics, the health profiles of each social position, and the health differences between social positions, will be affected. Hypothetical models show that mobility may, in theory, generate considerable differences. Using such models, Stern (1983) demonstrated that all the health differences between social classes could be accounted for by mobility. But hypothetical models are of course no proof in themselves, and a main theme has been to what extent empirical evidence indicates that social mobility is important in generating the typical health inequality patterns.

Social mobility explanations vary considerably. To illustrate this, I will give two examples.

8.1.2. Himsworth: intergenerational mobility and genetic selection

Himsworth (1984) is known for advancing a hypothesis which explains the continuing health inequalities by social variations in *genetic endowment*. Among health sociologists his paper is mostly ignored (Macintyre 1997: 732). His account is nevertheless referred here as it represents an extreme variant of the social mobility explanation of health inequalities.

Himsworth's starting point is that although infant mortality in England has decreased enormously, the relative differences in infant mortality between the occupational classes have been astonishingly constant. Improvements in

standards of living and the availability of health services explain the general absolute decline. But why is the same social pattern in infant mortality differences repeated, from each generation to the next?

Himsworth believes that the crucial factors are people's intellectual, physical, and reproductive capabilities. These capabilities are, according to Himsworth, not simply produced by social and physical environments. They are polygenetically determined by favourable or unfavourable combinations of genes. They are moreover related to each other. Persons with higher intelligence (i.e., a genetically determined potential for education and rewarding jobs) have usually better physical capabilities (better health genes), and in addition also better reproductive capabilities (genes favouring the ability to give birth to healthy children).

From these assumptions, Himsworth constructs an explanation for the continuing social differences in infant mortality (and by implication, also for other health inequalities between occupational classes). Human societies "exist in a state of flux". People move up and down on the social scale, because of educational attainment, occupational careers, and marriage. These mobility processes have as their results that people attain a level in society which corresponds to their genetic endowment. Those who are downwardly mobile "sink to a social level which accords better with their intellectual, physical and reproductive capabilities." The upwardly mobile rise "to social levels which accord more closely with their personal characteristics" (Himsworth 1984: 169). Therefore, people in the lower strata have less intelligence and more ill health, and low status mothers have less reproductive capabilities (i.e., give birth to more unhealthy infants) and less "maternal competence" (less ability to take care of their children). The opposite happens among higher strata: the higher on the social scale, the higher intelligence, the better health potential, and the better reproductive capabilities and maternal competence.

Although the distribution of genetic endowment generally follows the social hierarchy, this does not imply that children born into low status families always have less good genes than children from the higher strata. "Dull parents occasionally have a brilliant child" (p.171), brilliant parents may get dull children, and these deviating children will, because of mobility processes, be relocated in the social hierarchy so that they attain social positions in accordance with their personal traits. In this way Himsworth explains the continuing health differences between occupational classes.

I will suggest that his hypothesis is largely contradicted by research findings of the environmental influences on health, reproductive capabilities, and mobility patterns. A crucial point is of course what human genes are, how

they vary between individuals and, more important, between *social* strata, and what they determine. This is, to the present day, largely an unknown area (see below section 8.3.1.). That genes favouring intelligence, health potential, and reproductive capabilities correlate at the individual level, is one problematic point which Himsworth has slight evidence of. Another problematic assumption is that intellectual, physical and reproductive capabilities are primarily genetically determined, so that the influence from genes overshadows the differences in social and material environments. That personal traits (which even are genetically, and not socially, produced) determine social mobility is a third speculation, hardly consistent with mobility studies which point out the importance of social background, educational opportunities, and structural change for status attainment.

Himsworth's hypothesis is perhaps best characterised as a *speculation* that people's genetic endowment is behind whatever social pattern and social process we observe. This weakness, I suggest, is the main reason why his contribution is neglected, and not primarily the embarrassed feelings of sociologists when confronted with genetic explanations of social conditions.

8.1.3. Mobility and changing patterns of health inequalities

A study of mobility effects on health inequalities of a rather different kind has been made by Dahl and Birkelund (1999). Data come from the Norwegian Surveys of Level of Living 1980 – 1995. Two tendencies are discovered in these materials. On the one hand, the (self-reported) health differentials between the occupational classes among those who are currently employed are becoming smaller. On the other hand, differentials between those employed and those without work (ages 25-66) are becoming more marked.

When looking only at the employed occupational categories, one therefore gets the impression that socioeconomic differences in health are declining. The authors discuss whether an equalisation of working conditions, or general improvements in the material standards for all social classes, or better health services, can account for this development. If so, the explanation is within the social causation paradigm. However, they reject this interpretation. Building on various other materials, they argue that the observations can largely be explained by developments in the labour market, together with accompanying developments in the institutions which influence mobility patterns of employees into and out of employment.

The period under study has witnessed a relatively unstable labour market, sometimes with high levels of unemployment. Competition between capitalist firms has been intensified. The institutions of the welfare state providing

economic support for non-employed have developed. Early retirement and disability pension schemes have been expanded.

Due to these structural circumstances, various developments have taken place. The firms have become more selective as to the working capacity of their employees. Screening for health problems has become more effective when staff are hired, and failing health predisposes (openly or hiddenly) for being laid off when staff reductions are made. As the common chronic illnesses of the middle-aged and the elderly typically affect physical more than mental functioning, and physical strength is more important in manual and low status jobs, health selection through hiring and discharging is more effectively implemented for workers than for higher white collar employees. The structure of economic support for non-employed facilitates this. People may seek early retirement and disability pension, which, due to the wages/salaries differentials, appear more attractive the lower one is located in the income hierarchy.

Thus, those with failing health are "selected" out of employment, while better health increases the likelihood of staying, or becoming, employed. This mechanism acts more effectively among workers and low status occupations than higher up in the occupational hierarchy, and therefore the health differentials between the employed categories tend to decline, but at the same time the health differentials between employed and non-employed increase. Thus, mobility into/out of employment, conditioned by the overall situation in the labour market and implemented by firms and welfare state institutions, accounts for the changing patterns of social inequalities in health, and in particular for the decline in occupational health differences. "Paradoxically, class specific processes have generated a 'non-class' result, i.e., no class differences in health among the employed population" (Dahl and Birkelund 1999: 20). But the authors underline that social inequalities in health are not "really" disappearing. When the whole population, aged 25-66, are divided into occupational classes by including the non-employed in the categories they belonged to when employed, the health differentials between the occupational categories are of the same size in the 1990s as they were in the early 1980s.

8.1.4. Main themes

The examples given above illustrate that social mobility explanations for social inequalities in health are of rather varied types. In this chapter, I will concentrate on explanations of how *socioeconomic* health differentials are generated. Many themes are intermingled in these explanations, related to what types of health determinants are employed, to how the social mobility processes are understood, and to whether social mobility is seen as the main

process behind the health inequalities, or only as a contributory process interacting with social causation processes.

I will first describe findings as regards how socioeconomic inequalities in health appear in different age groups. Thereafter, I will discuss hypotheses about how a health potential, and especially *social variations* in this health potential, may be established early in life, with longterm effects on later health in adulthood. After that, I will address hypotheses about how social mobility may be involved in creating social variations in health at different stages in people's life courses. Finally, I will discuss whether research indicates that social mobility is *important* in generating socioeconomic differences in health.

8.2. Health inequalities in different age groups

For obvious reasons, those engaged in discussions about the role of social mobility in creating social inequalities in health will be particularly interested in how socioeconomic health differentials appear at different stages in life. Generally, social mobility explanations take a longitudinal view and ask whether the pattern of health inequalities at one time point is related, due to mobility processes, to the health variations at an earlier time point. Thus, social mobility explanations, as well as explanations which reject that mobility plays any important role, will usually refer to empirical findings about the health variations in different age groups – in infancy and childhood, in adolescence, among young adults, in middle age, and among the elderly. These empirical findings will be surveyed below.

Most studies indicate that there are social variations in health already among infants and children, when they are classified according to their social background, i.e., their parents' social status. As to mortality, evidence from various countries clearly indicates socioeconomic gradients among the youngest (see for instance Arntzen et al. 1988, Vågerö 1992, Bakketeig et al. 1993, Robert and House 2000). Mortality is however very low in these age groups. As to chronic disease and other types of illnesses, the evidence is not unambiguous. Various studies find some, sometimes small (Grøtvedt and Belsby 1995, Kohler 1990, West 1988), sometimes more marked (Wadsworth 1986, Finnvold and Nordhagen 1996, Gissler et al. 1998, Halldorsson et al. 1999), differences in ill health among children. Moreover, it is often supposed that there are significant socioeconomic differences in children's health potential, i.e., bodily or mental traits which are not, in themselves, illnesses or medical disorders, but which is supposed to influence the likelihood of later disease and mortality. How such a health potential may be established, and to what extent it determines later health, will be discussed in the next section.

Among adolescents, serious somatic diseases, disabilities, and mortality are rather rare. An extensive literature review of West (1997) suggests that among teenagers, there are small manifest health differences according to their social background. Thus, it may be that socioeconomic differentials in health are relatively lacking among youngsters. Whether this is so, is however disputed (West 1988, Blane et al. 1993a: 5, Macintyre 1997: 738, Bartley et al. 1997, Halldorsson et a. 1999). It has been suggested that the observed lack of social class differentials among teenagers only appears when ordinary health indicators (longstanding disease, disabilities, acute illnesses, mortality) are measured. Other, not-so-visible indicators about psychological health or health potential may perhaps show larger social differences. Given that there are small social variations in health in this age group, there may nevertheless be significant individual variations.

Generally, in young adulthood, the socioeconomic health differentials are quite clear, and they persist not only into middle age (where the empirical evidence is overwhelming), but also among the elderly (Dahl and Birkelund 1997, Arber and Cooper 1999). Whether socioeconomic health differentials *widen* with increasing age has been examined in particular. Different measurement methods (absolute versus relative differences, for instance) may give diverging answers. It has been suggested that the socioeconomic gradients in health are steeper among the middle aged than among younger adults and the elderly (European Science Foundation 2000, Robert and House 2000), but also that socioeconomic health differentials are of more or less the same magnitude in all adult age groups (Power et al. 1996a, Bartley and Plewis 1997).

Thus, although the evidence does not agree completely, it seems that there are significant socioeconomic health differentials "from the cradle to the grave" (Vagerö 1992). Disagreements concern in particular whether these inequalities are smaller in adolescence, and moreover whether their magnitude is particularly large among persons aged about 40 to 60 years. These findings raise several questions. One of them concerns the impact of childhood health or health potential on later, adult health. Several social mobility explanations operate with a notion of health determinants which imply that the roots of adult health are found in childhood. These hypotheses will be discussed below.

8.3. Health potential as a health determinant

As said above, the inclination of social mobility explanations is often to assume that a significant health potential has been established at some early point in life, and thereafter continues to have an overriding influence on the

subsequent health status. In this section I will discuss some approaches to this understanding, in terms of genetics and social and biological programming. Thereafter I will address the question about what an early established health potential means: does it *determine* later health?

8.3.1. The genetic health potential

The genetic endowment of humans is, in a very essential way, the potential upon which the health of humans depend. Basically, it is the genes which distinguishes between humans and other species, which determine what kind of nutrition which is necessary for human survival, what kind of heat, cold, and air that are compatible with human existence, etc.

Thus, genes constitute *the* fundamental health potential. But health is seldom based on genes alone. Health evolves within particular environments. As to the relationship between genes, environments, and health, Baird (1994: 134) makes a three-fold distinction. First, there are environments which are so adverse that health damages follow whatever genes one is equipped with, second, there are genes which are so harmful that health disorders emerge however favourable the environments, and third, there are disorders which develop through some kind of interaction between the genetic endowment and the environment.

There are environments which are so hostile that no individual is able to adapt without health damages. Grossly inadequate diet or toxins are incompatible with health, whatever genes one has. On the other hand, an individual may be equipped with extremely "unhealthy" genes. Thus, either because of genes inherited from parents, or because of mutant genes, the individual will develop health disorders within any normal environment. Such health disorders are often termed single-gene diseases (Conrad 1999, Martin 1999) or wholly genetic disorders (Wingate 1979: 194-195). Medical science knows of a large number of such wholly genetic disorders: cystic fibrosis, haemophilia, Huntington disease, thalassemia, etc. These diseases are however very rare and constitute only a minor part of the ill health panorama.

Thus, as regards population health, the wholly genetic disorders are of little interest. More important for the health panorama is the third type of disorders: those which are assumed to develop through the interaction between genes and environments. For a large number of acquired disorders, among them the major killers of today (heart disease and cancer), a number of common chronic diseases (diabetes, schizophrenia, rheumatoid arthritis), and some infectious diseases (e.g. tuberculosis), it is supposed that hereditary

factors also play a role. This implies not only that genes and environments interact. Generally, health and ill health are produced by this interaction, and in this sense the potential for health is always hereditary. When heredity is said to be involved, the precise meaning is that genes are supposed to *vary* in a way which equip some people with a better potential than others for withstanding unfavourable environments and profiting from beneficial environments.

Generally, it is believed that there are such variations between individuals relevant for the common health disorders. However, many uncertainties about this exist. Epidemiology has shown how practically every common health disorder varies with environmental factors, but also that some disorders tend to be more frequent in some families than in others, perhaps indicating significant variations in the inherited genes. A large international project – The Human Genome Project – is currently trying to map all human genes (Conrad and Gabe 1999: 505). From this, information about how genes vary and how they are related to health may emerge, but to what extent remains to be seen.

Thus, as regards the major and common health disorders, variations in genes may be important, but definite knowledge is often lacking and generally difficult to obtain. There are large methodological problems involved when one tries to distinguish between the effects of heredity and environment. Estimations of the role of heredity vary considerably. Tarlov suggests that "The best that can be done with the current state of knowledge is to assume tentatively that genes as a determinant of health account for 1-5 per cent of the total diseases burden of man" (Tarlov 1996: 73). Others allow for a larger role of genes. To answer "how much genes determine" is moreover complicated because it depends on the degree of variation, not only in genes but also in environments. Whether diseases appear as hereditary depend on the variation in environmental influences. As pointed out by Rose (1985): if environments were equal, any variation in disease incidence would be due to genetic differences, but if environments vary substantially, the role of genes will tend to disappear compared to the environmental influences.

In the context of this thesis, the central question is whether genetic variations are of any relevance for the emergence of *social* inequalities in health. This entails not only that there are significant individual variations in the genetically determined health potential, but also that these variations are socially patterned, i.e., systematically associated with major social structures. Wilkinson claims that no research has been able to demonstrate social inequalities in genes relevant for the susceptibility to ill health. He adds that the likelihood is rather that the lower classes, because of more disadvantageous environments, will develop better genetic resistance than the

higher classes: "The increased burden of illness from environmental sources would tend to increase the scale of genetic selection for the ability to resist various diseases. Thus, it is quite possible that survival of the fittest provides a stronger selective filter for good genes in lower classes" (Wilkinson 1996: 61). Baird draws the same main conclusion as Wilkinson, albeit with other arguments: "There is significant social mobility and intermarriage, enough that over several generations the gene distribution would be similar in the different social classes" (1994: 143).

Thus, today the hypothesis that social inequalities in health are a result of a more adverse genetic health potential the lower people are located in the social hierarchy, is speculative. Accordingly, apart from Himsworth and a few others, social mobility explanations of social inequalities in health do not rely on the idea that a genetically determined health potential varies significantly between socioeconomic social positions.

8.3.2. Psychosocial and biological programming

The research field has therefore been more interested in how a potential for later health may be established in the very early years, by means of other mechanisms than genes: i.e., by psychosocial or biological programming.

Psychosocial programming has been previously discussed in terms of attachment theory (section 7.3.1). In general, the focus is on psychosocial surroundings in early life, conditions of love, care, security, stability, attachment patterns, etc. When such circumstances are deficient, the child's psychosocial defence mechanisms may be lastingly hurt, and he or she enters the world of adolescence and adulthood with less capability to handle stressors and a higher vulnerability to disease. Thus, the child is programmed to be more susceptible to future health problems. Given that the ability of families to foster a psychosocially health potential varies with social position, less coping capabilities and higher susceptibility to later health problems may emerge in children brought up in disadvantageous social surroundings. Included in such a health potential may also be behavioural components: feelings of hopelessness and lack of self-confidence may bring about a higher probability of engaging in unhealthy behaviours when reaching adolescence and adulthood. Different terms are used for such early psychosocial programming, for instance the social imprint hypothesis (Bäckman and Palme 1998) or social programming (Vågerö and Illsley 1995). The latter concept sums up "those influences in early life which directly or indirectly determine adult health or which interact with adult experience to determine adult health" (p. 231), and may include not only psychosocial but also material circumstances.

Biological programming (also the biological imprint hypothesis) is particularly associated with the research of Barker and colleagues (1992), but similar hypotheses have existed for a long time (Kuh and Davey Smith 1993), and were for instance advanced by the Norwegian researcher Anders Forsdahl during the 1970s (Forsdahl 1977). The essence of this hypothesis is that the prenatal phase, before birth, but also infancy and the very early years, are important stages for the development of the human organism. The physical environments encountered in this phase influence the strength of various bodily systems. Biological programming has been defined as the process "whereby a stimulus or damage during a sensitive period in the development of the foetus result in a permanent injury of the structure and functioning of the organism" (Nordhagen and Bakketeig 1994: 3570).

Thus, nutrition deficits in pregnancy, infancy and the early years, or other physical influences from, for instance, parents' smoking, may result in a sub-optimal growth of the foetus and infant, and to impaired developments of various organs. Such injuries may lastingly determine bodily functions. They may lead to "irreversible consequences in the arterial structure, and increase the risk of high blood pressure later in life", to "impaired glucose tolerance" making diabetes more probable, and to deficient development of the respiratory organs increasing the likelihood of respiratory diseases (Wadsworth 1997: 861, Kuh and Davey Smith 1993: 117-118, Nordhagen and Bakketeig 1994: 3573).

To indicate such sub-optimal growth, Barker and associates often use indicators such as placental weight, premature births, birth weight, head circumference, length of the newborn, and growth during infancy. They have been able to show that such circumstances of the newborn influence the probability of various adult diseases and premature mortality (Barker 1992). Thus, the general hypothesis is that the health potential is significantly influenced by material conditions during early (including prenatal) life, and adult health is more or less biologically programmed already in the child.

Social mobility explanations of social inequalities in health often make use of notions about such an early health potential, produced by psychosocial or biological programming. The implication is usually that there are social variations in health potential among children and adolescents. The objective is then to examine how these social differences in health among children are transformed, due to various types of social mobility processes, into socioeconomic health inequalities among the adult population. However, when psychosocial or biological programming hypotheses are utilised, the initial social differences in health in childhood are usually seen as a result of *social*

causation. Deficits in the psychosocial or biological programming are seen as a result of decreasing psychosocial and material resources the lower on the social scale the families of origin are located. Those who are born into families with low material standards, less resources, and poor social relations, will establish more deficient resistances to environmental pressures and hazards, and will be more susceptible to future health failures. Thus, social mobility explanations for health inequalities will not seldom also refer to social causation processes when explaining the social variations in the early established health potential.

8.3.3. Past or present: which is more important?

A next question is: to what extent is adult health *determined* by the influences from early psychosocial and material conditions? An important theme in health inequality research during the last decades has been whether the early established health potential has an overriding influence on later, adult, health, or whether differences in environments during adult life also are significant and maybe even dominant.

Generally, the answer to this question is: it depends. If social variations in adult environments are large and variations in health potential small, the adult environments will be the major factor. If differences in adult environments are small and differences in health potential large, the health potential will dominate health in later life. In theory, both alternatives are possible. Research has to take as givens the actual circumstances, and has examined whether the early established health potential or the variations in adult environment play the major role, under the social conditions prevailing in present-day societies,

As Rahkonen et al. (1997a) ask, which is the stronger determinant of adult health – *the past or the present*? Numerous studies have addressed this question, using data on childhood circumstances (birth weight, childhood health, social class in childhood, material deprivation, familial conflicts) and on adult conditions (adult health, adult mortality, social class in adulthood). Often, survey data with retrospective information are used (Lundberg 1991, 1993, 1997, Lynch et al. 1994, Peck 1994, Rahkonen et al. 1997a, 1997b, Mheen et al. 1998). There are also prospective studies (Wadsworth and Kuh 1997, Bäckman and Palme 1998), studies which combine retrospective and prospective data (Blane et al. 1996b, Hart et al. 1998, Davey Smith et al. 1997, 1998), and studies using ecological data (Ben-Shlomo and Davey Smith 1991).

The results are not unanimous. However, the overall impression is that childhood circumstances cannot explain away the effects of adult conditions,

but neither can adult circumstances explain away the lasting effects of disadvantageous conditions in childhood and adolescence. Studies differ however to the extent that they declare childhood circumstances or adult conditions as the stronger determinant. The major conclusion is nevertheless that the idea that early life provides a health potential which continues to act on later health during the entire life course has substantial credibility, but the health potential is not the only factor, and later circumstances influence the health outcomes considerably and can even overshadow the early potential.

8.4. Health variations and social mobility

Social mobility explanations for the emergence of health inequalities are associated with a major sociological theme: how are people moving between different social positions, and what are the reasons for these transitions? People start their lives within the circumstances of their family of origin. Through adolescence and early adulthood they typically attain an educational level and start their occupational career. In adulthood, their occupational trajectory develops, until some type of retirement because of old age or because of other circumstances. Likewise, they have a career within the marital structure, within the geographical structure, etc. Mobility may be intragenerational – social mobility within one's own life career – or intergenerational, meaning social mobility where the point of reference is the person's social background, for instance his/her father's or mother's occupational class or educational level. Central questions are what factors are crucial in making life courses different, and what types of conditions enhance or inhibit mobility transitions.

The study of social inequalities in health is related to this topic in terms of a special interest in how health is associated with people's mobility, and how this influences the patterns of health inequalities. How geographical mobility influences the geographical distribution of health conditions has been studied (Boyle et al. 1999), as has the question about marital dissolution and ill health: is the latter predisposing for later divorce, or a consequence of divorce (Mastekaasa 1993a: 13-19, 1993b)? The main interest is however how health variations in the social class, occupational and labour market structures are associated with mobility. In the following, I will outline three themes which in particular are examined by researchers discussing social mobility explanations.

8.4.1. Early health potential and social mobility

Socioeconomic differences in children's and youth's health and probably in their health potential, created by psychosocial or biological programming, are

often a premise for social mobility explanations. Usually, these differences are considered a result of social causation, but it is also sometimes suggested that also mobility may be involved, in terms of the parents' social mobility.

Illsley's research shows for instance that mothers who have been downwardly mobile, i.e., marrying into social positions lower on the social scale than their fathers' positions, tend to give birth to children who are less healthy than the children of mothers from the same social background who are not downwardly mobile (Illsley 1955, 1986). The explanation may be that downwardly mobile mothers have less capacity for generating sound psychosocial or biological programming. These results were also utilised by Himsworth (1984) in his efforts to argue that the downwardly mothers have usually an inferior genetic equipment which they pass on to their offspring.

Mobility may also be involved in a more peculiar way. It has for instance been suggested that children's illness influences their parents' career. Thus, low income, more downward mobility, and less upward mobility, will occur more frequently among the parents of ill children. Thereby, an association between the parents' status and the children's health is produced. That such processes occur very often are usually rejected (e.g. Blane et al. 1993a: 3).

Whether social mobility has produced social variations in children's health (or health potential) is however a minor theme, compared to the question about how social variations in children's health are transformed into social inequalities in health when the children become adults. A familiar result of mobility studies is that people have a tendency of attaining a location in the socioeconomic structure which is fairly close to their parents' (or fathers') location. Worker's children tend to become workers, white collar children enter white collar occupations, and sons of farmers become farmers. Although upward social mobility is frequent, much of this mobility is *structural*, generated by overall changes in the occupational structure (Ringdal 1994, Eriksson and Goldthorpe 1992). Often, people attain a nominally higher occupational position than their father's without having attained a higher *relative* position in the social hierarchy. That the class structure has a tendency to reproduce itself in this way may have nothing to do with health. It may follow only from constraints and opportunities as regards economic, social and cultural capital, which differ between different social backgrounds.

Given that mobility is usually of this type – i.e., little "real" inter-generational vertical mobility – this may partly explain why socioeconomic health inequalities emerge in adulthood. If children's health potential varies with parent's occupational class, and if adult health conditions are strongly formed by

childhood health potential, the result of such mobility regimes will be social inequalities in health in adulthood. Thus, adult health inequalities may emerge in consequence of the health inequalities already present in children and the tendency that adult health depends on early health conditions, in combination with intergenerational mobility patterns which tend to locate the offspring at levels in the social hierarchy corresponding to the location of their parents.

8.4.2. The transition from adolescence to adulthood

The topic which perhaps has attracted most interest is whether health among adolescents is associated with educational attainment and the occupational career they start in young adulthood. Thus, it has been suggested that young people with poor health or poor health potential, or with other characteristics which predispose for later health failures, tend to obtain lower educational levels and to enter occupations located in the lower end of the occupational hierarchy. Mobility in the "crucial period in the life-course ... between childhood, adolescence and early adulthood" (West 1991: 380) may be health-related, and therefore create social inequalities in health in adulthood. Often, this explanation has been connected to hypotheses about small social class variations in health in adolescence, followed by marked social inequalities in health in early adulthood (West 1997, Blane et al. 1993a: 5-7). In itself, this explanation is however independent of the extent of social variations in health among youth. The central question is how individual variations in adolescent health predict later social positions in the occupational structure.

Health problems may directly hinder educational careers. Prolonged and frequent sickness absence, or disabilities (deafness, physical disabilities) may make it impossible or very troublesome to perform in educational institutions. Adolescents may choose "easier" educational options because of their health problems, or their disorders will obstruct and impede educational success. As to psychological problems among adolescents, for instance, Miech et al. (1999) find that conduct disorders and attention deficit disorders, but not anxiety and depression, predict educational failure. The reasons may be that conduct and attention disorders, but not depression/anxiety, are particularly disqualifying for educational attainment. The organisation of educational institutions is important, i.e., whether institutions are equipped with, or willing to provide, the special requirements of young people with disabilities and various health problems (cf. Grue 1995).

Ill health in childhood has been found to contribute to less educational and occupational achievement in early adulthood, even when controlling for

the children's social background (e.g. Wadsworth 1986). Manifest health problems will often act as a reason for choosing less demanding and often low paid occupational careers. More important, employers will regularly discriminate against young people with health problems (West 1991), either denying them employment, or preferring the healthy for the more attractive jobs. Thus, the more unhealthy may be channelled to jobs where the competition for work is less intense, i.e., to lower paid and less attractive jobs.

These are examples of *direct health selection*, where manifest health problems contribute in how young people are distributed in the educational structure, and thereafter in the occupational structure as young adults. The consequences of such selection processes will be a higher frequency of young adults with manifest health problems in the lower ends of the occupational hierarchies.

However, manifest and disabling health problems occur rather seldom in these age groups, and the overall effect of direct health selection in the transition from adolescence to adulthood for the production of socioeconomic health differentials is often believed to be slight. Processes of *indirect health selection* – health-related mobility where manifest health problems are not directly a contributing cause – have attracted more attention (Blane et al. 1993a). A favourite in this research is body height (West 1991: 381-382, Macintyre and West 1991, Peck and Vågerö 1989). When controlling for other relevant circumstances, taller people are regularly found to be more successful in educational and occupational careers. Being taller than the average may predispose adolescents to be more self-confident and more ambitious. They may be considered more attractive and therefore preferred by employers and by educational institutions. If height and health potential are positively correlated, higher social positions will be more populated by healthy people due to the social processes which select taller young people into the better positions.

That height correlates with better health and lower mortality has been demonstrated in many data (e.g., Waaler 1984, Peck and Vagerö 1989, Silventoinen et al. 1998). The reasons are however unclear. Tallness may indicate a better health potential, but the association may also emerge because of the better environments experienced by taller people in childhood (better nutrition leads to tallness), or during adulthood because of their higher tendency to upward mobility (West 1991: 378, Dahl 1996). Selection for height may be an example of a more widespread tendency of selecting for *attractiveness*, where also other bodily criteria than height may be relevant,

such as weight and body-mass (Macintyre and West 1991). These characteristics may also constitute a health potential, similar to height.

Indirect health selection may even take other forms. Karvonen et al. (1999), for instance, have showed that, independent of social background, those adolescents who displayed more unhealthy behaviours when aged 16-18 were less likely to be upwardly mobile and more likely to be downwardly mobile. Thus, it seems that unhealthy habits established in young age, which may increase the chance of later health failures, were also predictive of their later locations in the occupational hierarchy. Similar findings have been reported from other studies (Glendinning et al. 1994, Blane et al. 1993a: 7-8). Indirect health selection may also be related to social background. Social background may influence mobility chances (higher background predicts higher educational attainment, which predicts more successful occupational careers), and social background may influence the health potential (higher social background is linked to better nutrition, better hygiene, and better access to health services in childhood). Thus, through the mechanism of choosing better educationally qualified adolescents for better occupational careers, one is also inadvertently choosing people with better health potentials, and an association between higher positions and better health will be generated.

8.4.3. Health and mobility during adulthood

Social mobility explanations focussing on adulthood are usually less oriented towards using an early established health potential as part of their arguments. More often, the focus is simply that adult people may develop health problems, for whatever reason – perhaps because of the environments they have encountered – and the question is what consequence this has for their careers.

During adulthood, the most prominent health-related mobility in our societies seems to be selection *out of* employment because of ill health. Regularly, it is found that health problems dispose for longterm unemployment, transitions to early retirement or disability pensions, or (for women) returning to homework (Dahl and Birkelund 1999, Mheen et al. 1999, Bartley and Plewis 1997, Claussen 1999). Such health selection will often be of the direct type, implying discrimination and exclusion by employers, but also self-selection because of the troubles people feel in combining their health problems with employment.

To what degree health is related to successful or unsuccessful occupational careers in adulthood is more disputed. The *drift* hypothesis has been a longstanding topic in American research. Often it refers to intergenerational

mobility, where failing health and disease – mental health and in particular schizophrenia are commonly in focus (Goldberg and Morrison 1963, Turner and Wagenfeld 1967, Mechanic 1978: 217, Fox 1990, Timms 1996, 1998) – are shown to contribute to downward mobility when compared with the social level of one's parents. But the drift hypothesis is also applied to intragenerational mobility, formulated, for instance, in the following way: "people who develop disabling chronic illnesses drift downward in social status over time because their disability prevents them from gaining and maintaining jobs up to their illness-free capacity. Thus, the lower strata become overloaded with people in poor health, not because of differential incidence of chronic illness, but because having a disability has a negative effect on socioeconomic status, thus ensuring that the chronically ill tend to move downward or remain in the lower strata" (Harkey et al. 1976: 194). Correspondingly, positive health, for instance in terms of physical or psychic strength, has been supposed to increase the likelihood of successful careers, a hypothesis which, as we saw above, was advanced by Sorokin.

Indications of health influences on adult occupational careers are found, but the effects are usually small (Blane et al. 1993a: 8-10). Transitions caused by ill health to part-time work may occur, leading often to less income but not necessarily to a decline in occupational level. Health-related transitions from worker to lower white collar occupations have also been found (Dahl 1993a, 1993b), reflecting difficulties in combining heavy manual labour with physical health problems. Some studies find associations between health indicators and the direction of intragenerational occupational mobility (e.g., Bartley and Plewis 1997), others do not (Rahkonen et al. 1997b, Mheen et al. 1998, 1999).

Accordingly, it seems that in contemporary societies, there are various institutional mechanisms which obstruct health selection, leading to drift downwards or to considerable upward mobility, from taking place to any large extent in the occupational careers during adulthood. Serious ill health leads primarily to exclusions from employment and into positions supported by welfare state benefits, not to employment at lower levels in the occupational hierarchy. Regulations of the rights of employees and the tight connection between education and occupational position will further impede drastic movements because of good or bad health, up as well as down, in the occupational structure. It may however be suggested that current research does not always have appropriate data for studying the details of possible health-related mobility. The frequent use of crude occupational classifications, and the lack of data on positive health, may lead to underrated estimations of what health variations means for differences in adult occupational careers.

8.5. Debates: The role of social mobility

8.5.1. The distinctive feature of social mobility explanations

Explanations for health inequalities *solely* by mobility processes, without social causation processes involved, are hard to find. That physical and social environments influence health, and that these environments vary considerably between social positions, are so well documented that hardly any explanation disclaims any role at all for social causation.

The distinctive feature of social mobility explanations is therefore that, within many processes of social causation, it is also claimed that there exist mobility trajectories which influence the generation of the typical patterns of social inequalities in health. Usually, the premise is that social mobility transitions are health-related, either directly or indirectly, so that there is a significant tendency that those who move upwards in the social hierarchy, or stay in the upper strata, are of better health or better health potential than those who move downwards in the social hierarchy, or stay at the bottom of it.

Usually, social causation explanations also acknowledge that mobility may sometimes be health-related, and that this may have some consequences for how health inequalities appear between social positions. The distinguishing feature of social causation explanations is that such health-related mobility is believed to be of very little significance for the overall pattern of health inequalities. Social inequalities in environments, rather than differences in mobility patterns for the healthy and the sick, are assumed to be the major explanation for social inequalities in health.

Thus, the two approaches coincide in that they acknowledge that health-related social mobility occurs, but they differ grossly as to *how much* they believe that social mobility is able to account for the health inequalities.

8.5.2. How much does social mobility explain?

From hypothetical modelling, it can – as mentioned earlier – be shown that the typical patterns of health inequalities can emerge solely from health-related mobility, with no social causation involved, given certain assumptions about the size of the different classes, the degree of mobility between them, and the degree of selectivity for health, i.e., how strongly social mobility co-varies with health. Although admitting that sufficient "information does not appear to exist currently", Stern (1983: 44) nevertheless argued that data indicate that the increase in mortality differentials in Britain could be interpreted as a consequence of health-related mobility, increasing mobility rates, and changes in the relative size of the occupational classes.

This was the topic of the exchange between Illsley (1986, 1987) and Wilkinson (1986a, 1986b, 1987) (see also section 4.4.4 above). Illsley's (1986, 1987) interpretation was that the disadvantageous classes were becoming smaller, and the advantageous classes were expanding. This change in the occupational structure implied both considerable intragenerational and even more intergenerational mobility. Some from the lower classes moved upwards in the occupational hierarchy within their own occupational careers, and even more of the children from lower status families entered occupational careers at a higher level than that of their fathers. This mobility, argued Illsley, was health-related. The upward movers were systematically of better health than stayers at the lower levels. The lower occupational classes shrank, and those who remained there were negatively "selected" because of health problems and therefore displayed particular high mortality rates.

There followed an "acrimonious round ... of the inequalities debate" involving "rather dubious arithmetic", according to West (1991: 377). The disputed point was how the coefficients indicating mobility rates and the selectivity for health in mobility transitions were related to the ensuing magnitude of mortality differentials. Here, disagreements were total. Wilkinson claimed that the contribution of health-related selection accounted for hardly more than a fifth of the overall health differences (1986b: 420-421). From the same data Illsley contended that health-related selection could have produced the major part of the observed inequalities (Illsley 1987: 221).

A complicated point is how differential mobility chances, associated with health variations, are translated into the overall patterns of health inequalities. Later contributions have tried to solve this question, from different angles. Joffe, for instance, analysed social variations in birth weight for babies born by mothers before the age of 23 (Joffe 1989). Using data on height and height-related social mobility through marriage, and making some assumptions about how mothers' height and birth weight could be expected to co-vary, he estimated that "16.3 % in the observed short-term gradient" (p. 618) in social differences in the babies' birth weight were due to the height-related mobility experiences of the mothers from fathers' to own social class.

Bartley and Plewis (1997) addressed another topic: the interpretation of age variations in health inequalities. Given that the environments during adulthood were particularly responsible for the development of health inequalities (the social causation explanation), one would expect that occupational differences in health would increase during adulthood, as the disadvantaged occupational positions, relative to the advantaged ones,

accumulate more and more health hazards over time. However, many data sets do not show such an increase in health differentials by age. Bartley and Plewis used data on occupational class differences in long-term illness in 1991 in England/Wales, and on previous intragenerational mobility 1971-1981, showing that upward mobility was to some extent associated with less health problems in 1991, and vice versa for downward mobility. In spite of this, health inequalities did not increase by age, but remained rather constant. This was explained by arguing that health-related mobility acted to *constrain* the expected development of larger health differentials. Even though social mobility was health-related, it restrained, rather than widened, socioeconomic health inequalities. The implication was thus that the patterns of social inequalities in health could not have been generated by health-related mobility. The hypothesis that social mobility constrains, rather than widens, social class health differentials was later supported by a study by Blane et al. (1999), and is also addressed by one of the studies included in this thesis (Elstad 2001).

Current research on mobility and its relationship with health inequalities provides therefore rather conflicting answers to the question about how much health-related mobility accounts for the existing patterns of health inequalities. Technical and measurement questions of a quite intricate nature are involved – are measurements of health available for all the relevant time points, how is inequality measured, and what methods are appropriate for calculating how differential mobility is transformed into patterns of health inequality? Apart from this, the difficulties are related to two main questions. First, what assumptions are made as to the role of social causation? In order to calculate the effects of mobility, which necessarily involve dynamic processes over time, estimations of the effects of the corresponding social causation processes are also required, in order to separate the effects of the two from each other (Elstad 2001). Second, what time period is investigated? The processes investigated – the parallel dynamics of health-related mobility and social causation – evolve constantly, and the choice of a particular time sequence may produce results which are not necessarily equivalent to the findings in data covering other time period, or longer time periods.

9. Five papers – content and location in the research field

9.1. The theme of the five papers

The preceding Chapters 4 - 8 have surveyed the main explanatory alternatives proposed by the research field for understanding how the patterns of social inequalities in health are generated. These are the artefactual explanation, three main variants of social causation explanations (materialist/structural, behavioural/lifestyles, and psychosocial), and social mobility explanations. The debates within the research field during the last decades can be seen as attempts to gauge the relevance of these approaches, in themselves, in conjunction with each other, and in confrontation with each other.

These explanations are the background for the large number of studies, addressing a very wide array of specific questions, which have appeared in the field. The five papers included in this thesis are contributions to the debates on some of the issues which are raised by these explanatory alternatives. These papers have been published in journals which are important forums for the international dissemination of studies on social inequalities in health. They participate, so to speak, together with numerous other studies, in the international accumulation of findings, knowledge, and discussion about these issues. The papers have not a single and narrowly delineated focus. They are, in a way, located at different *sites* in the research field. Paper I addresses the artefactual explanation. Paper II is oriented towards the concepts of the psychosocial explanation. Paper III is mainly directed to the understanding of trends, while Paper IV and V discuss how social causation and social mobility may interact in producing the patterns of social inequalities in health.

Thus, in one way or another, the five papers address the main theme: how are social inequalities in health generated? This is their common topic. Paper II is a theoretical and conceptual review of questions relevant for the psychosocial explanation. The four other papers are empirical studies, contributing with both theoretical discussions and empirical evidence.

The four empirical papers have some common characteristics. They use data from interview surveys conducted by Statistics Norway – the *Health Surveys* (1968, 1975, 1985, 1995) and the *Surveys of Level of Living* (1980, 1983, 1987, 1991). These surveys have data on health and social positions,

together with other information, for nationally representative samples of the Norwegian population. The details of these data, as regards sample sizes, response rates, and the variables employed, are discussed in each paper. The data are analysed by various statistical techniques, such as cross-tabulations, OLS regression, and loglinear analysis. Health information is obtained by *self-reports*, and a variety of health indicators – but of course not mortality – is analysed in these studies.

Moreover, the subjects of these four empirical papers are always in some way or another related to changes, trends, or some other types of dynamics developing over time. This is done by utilising the panel parts of the Surveys of Level of Living (Paper I, Paper IV), by constructing time series with data from subsequent surveys (Paper III), and by employing the retrospective information available in the 1995 Health Survey (Paper V).

In the following, I will summarise the papers, show their main results, and discuss how they are related to other studies within the field.

9.2. Summaries of the papers

9.2.1. Paper I: How large are the differences – really? Self-reported long-standing illness among working class and middle class men

This paper contributes to the investigation of the artefactual explanation. It is concerned with the measurement of health (cf. section 4.2.1). The specific topic is the interpretation of social class differentials from *self-reports* of longstanding illness: do such self-reports give biased measurements of the magnitude of social class health differentials?

Two hypotheses are raised. The first one suggests that disadvantaged social positions will have a higher propensity to report health complaints, compared to more advantaged social positions. This may be so because the social circumstances of disadvantaged positions produce lower thresholds for perceiving symptoms. Lower strata persons may react more violently to health threats. As they have less resources to cope with stressors (cf. section 7.3.1), the outbreak of health difficulties may appear more serious. The contrasting hypothesis suggests that advantaged social positions will be more inclined to note symptoms and report them. Their better circumstances may produce higher expectations as to what constitutes good health. They may be more observant of symptoms. Higher educational levels may imply that they are better prepared to interpret signs from their bodies, and closer relationships to the health services may lead to a higher level of diagnostication.

By means of an intensive statistical analysis of data from the Surveys of Level of Living 1980 – 1991, the characteristics of the longstanding illnesses reported by working class and middle class are compared. It is shown that workers reporting at least one longstanding illness have a higher average number of diagnoses than middle-class men. A higher proportion of workers with longstanding illness report that their illnesses limit their working capacity. The use of check-lists prompts the reporting of additional diagnoses more often among middle-class respondents than among workers. Using panel data, it is shown that a larger percentage of middle-class long-standing illnesses reported in the first wave of interviewing is not reported at the second wave. It is further shown that diagnoses commonly classified as less serious constitute a larger proportion of the diagnoses reported by middle-class men than by workers.

These findings support the second hypothesis outlined above. The longstanding illnesses reported by middle-class men seem to be less serious than those reported by workers. The conclusion is therefore that these ways of measuring morbidity are probably biased in the sense that they *underrate* the "true" magnitude of social class health differentials.

These findings are consistent with some, but not all, of the previous studies addressing this question. There are however not many similar studies focussing on reporting patterns in different socioeconomic positions. The general theme – whether social positions differ in their ways of reporting health complaints – has more often been addressed as regards other social structures. Studies comparing how men and women report health problems do not indicate that women's higher level of morbidity reflects lower pain thresholds among women (Bendelow 1993, Macintyre et al. 1999). On the other hand, many studies have found cultural and ethnic differences in how health problems are perceived and reported (Zola 1966, Koopman et al. 1984, Bates and Rankin-Hill 1994).

9.2.2. Paper II: The psychosocial perspective on social inequalities in health

Contrary to the other papers, this study has no empirical analyses, but presents an overview of diverse theoretical approaches related to the psychosocial explanation of health inequalities (cf. Chapter 7). It locates the current popularity of this explanation in the inability of both social mobility explanations and social causation explanations focussing on physical environments to explain satisfactorily how social inequalities in health emerge. A short comment on the etiological status of the psychosocial explanation is presented. Thereafter,

various lines within the psychosocial explanation are surveyed, in terms of *social stress research*, the *self-efficacy* model, the *sociology of emotions*, and the *social cohesion* approach. The tentative conclusion is that the psychosocial perspective represents an enrichment of the social causation explanations, rather than an alternative which can replace them. The contribution of this paper is to present a broad, albeit short, overview of different approaches within the psychosocial explanation. It argues that these approaches are interrelated and can together be seen as representing an important perspective within the research on social inequalities in health.

9.2.3. Paper III: Inequalities in health related to women's marital, parental, and employment status – a comparison between the early 70s and the late 80s, Norway

This paper addresses health inequalities between women in three social structures: the marital, the parental, and the employment structure. The topic is whether we can observe changes in the magnitude of health differences between these social positions (or combinations of these social positions) during the 1970s and 1980s in Norway, and the possible explanations for such changes. The overall hypothesis is that social change may have consequences for how the physical and psychosocial environments differ between social positions, and moreover that social change may lead to new social mobility patterns with consequences for how health inequalities among women appear.

Previous findings from many countries indicate that the health of married women is favourable in comparison with previously married, mothers are of better health than non-mothers, and employed women compare favourably to women without employment. The paper discusses the social causation processes which may generate these differentials, and suggests the relevance of both the physical and the psychosocial environments. It also notes that health-related mobility can be involved: ill health may predispose for not entering, or for leaving, paid work, and may restrict the number of births.

It is highlighted that these processes operate in a changing social context. New circumstances related to women's work have developed (e.g., a sharp rise in women's paid employment in this period, new types of working conditions, better possibilities for combining paid employment with household chores). Women's options for income have changed (more incomes from own work, developments in welfare state provisions). Women's marital positions change (increased number of divorces), and women's control over the number of births has been enlarged (increased availability of birth control methods and of legal abortions). These social changes may have modified the processes which

generate health inequalities. It is hypothesised that disparities related to marital status will diminish under the new circumstances (divorced women are now less stigmaticised and have better income options), disparities related to employment status will increase (for various reasons, particularly because of stronger selection mechanisms), while no particular reason is found for changing health differences related to parental status.

Using five subsequent cross-sectional surveys (1968, 1975, 1985, 1987 and 1991), the trends in health differences between social positions in these structures are examined. The hypothesis of increasing health differences between employed versus non-employed women is, by and large, supported by the results. As hypothesised, data indicate relatively unchanging differences related to parental status. The hypothesis of smaller differences in the marital structure is however *not* confirmed. The reasons for the unexpected result as regards differences in the marital structure (no change, rather than diminishing inequality) is proposed to be that the difficulties related to marital dissolution are primarily of a private character. Therefore, macro-social developments will have less significance for the processes which generate these health inequalities in the marital structure. – However, the data allow for no direct testing of the processes which are assumed to be involved. Trends are described, but the reasons for the observed developments cannot be tested directly, and are only discussed in terms of what seems probable and reasonable, given other evidence and knowledge about social developments in Norway in this period.

The conclusion of the paper is that the patterns of health inequalities among women seem surprisingly stable, on the background of the great changes which have occurred in women's lives in Norway during the 1970s and 1980s. The exception is the health differences related to the employment structure: differences between employed and non-employed women seem to have increased. This is also found in a later study (Elstad 1999a). It may very well be connected to mobility processes. Increasing employment rates of women may imply that those who remain non-employed tend to have particular high levels of ill health (cf. the processes described in section 8.1.3).

The general theme addressed by this study is how social change influences the processes which generate health inequalities. The health differences between men and women seem to have been rather stable during recent decades (Kjøller et al. 1995, Elstad 1999a, Lahelma et al. 1999). Also differences between socioeconomic positions do not seem to be immediately affected by relatively moderate types of social change, such as the restructuring of the labour market in recent decades in the Nordic countries (Kjøller et al. 1995, Dahl and Elstad 1999). Changes in levels of unemployment seem however to affect health-

related mobility out of employment and thereby influence the patterns of health inequalities (Dahl and Birkelund 1999, Lahelma et al. 2000).

9.2.4. Paper IV: Employment status and women's health – exploring the dynamics

In line with findings from various other countries, this paper shows that in Norway during the 1980s, self-reported health problems (limiting long-standing illness and psychological problems) varied systematically according to women's position in the employment structure. Full-time employed women had relatively few problems, part-time employed somewhat more, there were even more among home-makers, and women receiving pensions or welfare state benefits reported most health complaints.

How are these health differences generated? The paper tries to examine the role of both social causation processes and social mobility processes, in particular health-related mobility in the transition from adolescence to adulthood (see section 8.4.2) and during the adult lifecourses of women (see section 8.4.3). The social causation hypothesis proposes that less favourable social and material circumstances produce more health problems. Compared to the full-time employed, the part-time employed and even more homemakers and recipients of welfare benefits may suffer from effects of low income, low status and prestige, and less satisfying daily lives. The mobility interpretation is that poor health may lead to disadvantaged positions, and the more health problems, the more *distant* positions from full-time employment will be occupied, because of processes of direct or indirect health selection.

Using the panel components of the Norwegian Surveys of Level of Living 1980-1991, the paper demonstrates that better health in youth predicts full-time employment in early adulthood, while less good health in adolescence is more often followed by part-time employment or homemaking. Furthermore, the paper shows that during adult life, numerous changes take place, not only in employment positions, but also in health status. The typical pattern of health inequalities, at every time point, is therefore a continuously produced result of changes in both employment position and health. The dynamics of these changes are investigated. Using loglinear models it is tried to understand how health at one time point influences stability or change in employment positions during the following time period (i.e., health-related mobility), and correspondingly how employment positions at the first time point influences health developments in the following years (i.e., health changes produced by social causation processes).

Patterns reflecting both social causation and health-related mobility processes are discovered. Thus, inequalities seem to be generated by a mixture of causation and mobility. The results indicate that health-related mobility is frequent as regards women's movements between employment positions. It appears that health-related mobility is quite important as to how the health inequalities in this particular social structure are produced, but also social causation processes are indicated by the results. The general conclusion is that social inequalities in health may be generated by complex combinations of simultaneous social causation and social mobility processes.

9.2.5. Paper V: Health-related mobility, health inequalities, and gradient constraint: Discussion and results from a Norwegian study

The starting point of this paper is recently published studies which argue that the role of health-related mobility is primarily to restrict and diminish the magnitude of health differentials (cf. section 8.5.2). This is termed gradient constraint. This contrasts the widely held view that health-related mobility will always tend to increase health gradients – if not much, the effect is at least supposed to be in this direction. If mobility leads to gradient constraint, this implies that social inequalities in health are not produced by health-related mobility, and thus that social mobility cannot play any important role in generating the typical patterns of social inequalities in health.

The paper examines this hypothesis both from a theoretical and an empirical angle. Using hypothetical models, it is shown that when health-related mobility has a characteristic structure, its effect is to make health differentials smaller. This occurs typically when the initial health differences are relatively large, social mobility is widespread, and the selectivity for health (i.e., how closely mobility chances co-varies with health variations) is relatively small. Thus, gradient constraint may occur, given particular circumstances. But it is also shown that health differentials may widen in consequence of health-related mobility, typically when mobility chances vary closely with health. From hypothetical models, it is therefore shown that health-related mobility may both act to widen and to reduce social class health differentials, and the concrete result depends on several circumstances, among which the selectivity for health in the mobility transitions is particularly important. Thus, the paper supports the idea of gradient constraint put forward by Bartley and Plewis (1997) and Blane et al. (1999), but it disagrees with these former papers as regards their ways of explaining the mechanisms of gradient constraint.

The empirical part of the paper examines materials from the Norwegian Health Survey 1995, with data on male respondents' own occupational class,

fathers' occupational class, own bodily height, and some other health indicators. It is shown that social mobility seems to be health-related. Upward intergenerational mobility is linked to better health and downward mobility to worse health. Social class differentials are at about the same level, both when the sample is classified according to their fathers' occupational class and when the sample is classified according to their own occupational class.

The discussion asks what these empirical patterns indicate as regards the role of health-related mobility in producing the magnitude of health differentials. It is proposed that as health-related mobility processes and social causation processes occur simultaneously, it is difficult to estimate how health-related mobility has influenced the size of the health differentials without knowing how the sample's health has been affected by social causation processes (differential exposure to health hazards). In this study, the role of social mobility is problematic to estimate, because health is measured in the post-mobility situation and data on respondents' health before mobility are lacking.

Thus, as regards the health indicators which probably are influenced by environments during the respondents' lives, the data do not enable a precise estimate of how health-related mobility has influenced the size of the health differentials. This is however possible as regards one of the indicators, *height*. The special characteristic of height is that it is not influenced by adult circumstances. Therefore, changes in height differentials from the pre-mobility to the post-mobility situation are produced solely by height-related social mobility. Using this line of reasoning, the conclusion seems to be that height-related mobility has redistributed these male respondents in a way which has reproduced the occupational class differentials, from class of origin to class of destination, at about the same level. Thus, gradient constraint, i.e., social mobility which reduces the occupational class differentials, seems not to have occurred as regards the height differential in this data material.

The general conclusion of the paper, similar to the conclusion drawn from Paper IV, is that the pattern of social inequalities in health at any particular point in time will often be produced by intermingled social causation and health-related mobility processes. Thus, it is suggested that the opposition between the social causation and the social mobility approaches to how health inequalities are generated, should be replaced by a more integrated view about the combined effects of these processes. To show the specific role of each of these processes requires data which, unfortunately, often are not available.

10. Concluding discussion: towards an integration of the explanations

10.1. The state of the research field

The preceding chapters have delineated the research field of social inequalities in health and surveyed the main explanatory alternatives – their views on social processes and health determinants, their main empirical themes, and some of the discussions and controversies which are related to them. The purpose was, as said in Chapter 1, to make a bird's-eye view of the research field. The papers which are included in this thesis have been summarised in Chapter 9. I have tried to show how they contribute, to some extent, to the clarification of the main topic – how social inequalities in health can be explained. These studies are a small, but hopefully not trivial, addition to the already very large number of other studies published by researchers within this area.

What are the results of these endeavours? Has there been *progress*? Is the research field now able to produce a comprehensive and convincing explanation of how social inequalities in health are generated?

In this final chapter, I will discuss this question. I will propose that what we have seen is an enlargement of knowledge, accompanied by a growth of explanatory alternatives. The research field has been productive. However, its efforts do not seem to lead towards a narrow concentration on one, or a few, explanations. Rather, one can observe an increasing number of "tools" which seem relevant. This points towards an *integration* of the explanations, and the last part of this chapter will outline two such integrative attempts.

10.1.1. Developments in width and breadth

An optimistic view of social research is as follows. An issue – a social problem, a social condition, a social pattern – is discovered. How is this social condition generated? Research starts with loose, approximate ideas. Gradually, these ideas become more clearly formulated. Hypotheses are formulated and tested. Some are rejected, some are confirmed, as to some decisions are hard to make. Through this process, the number of credible hypotheses are reduced. The focus becomes more narrow. The surviving explanatory alternatives are made more precise, and joined together. Fruitful hypotheses are integrated into a consistent framework. Research advances towards an adequate explanation, by abandoning the implausible hypotheses and refining the plausible ones.

Is this metaphor consistent with what has occurred within the research field of social inequalities in health?

Certainly, we observe developments. By and large, the materialist/ structural explanation was the main focus during the first century of capitalist industrialisation. The epidemiological transition and rising levels of living in the West after the 2nd World War shifted attention to the behavioural/lifestyles explanation. Towards the end of the 20th century, the psychosocial perspective came more in focus. Alongside these social causation understandings, social mobility explanations have always existed as challenging alternatives.

In my view, it is doubtful that these developments imply, primarily, that unfruitful hypotheses have been eliminated, so that we are left with a more narrow set of credible propositions. The materialist/structural explanation is now widely seen as insufficient. But it is not obsolete, and often new evidence appears which indicates its continuing relevance. The behavioural/lifestyles explanation, which seemed to gain supremacy for some time, is hardly able to account fully for these health inequalities. As to the psychosocial explanation, it seems premature to expect that it will entirely replace the former approaches. Social mobility explanations continue to present a relevant opposition to the social causation paradigm. Few claim that mobility accounts for the bulk of health inequalities, but its relevance is nevertheless widely accepted.

Accordingly, hypotheses founded in all the main explanations and their subvariants are regularly put forward and examined by empirical studies. Many of these hypotheses get support, but they are seldom confirmed beyond doubt. New hypotheses are launched, which again are often partly confirmed, but also raise new doubts. The research field has, on the one hand, often been unable to discredit old explanations in a definite manner. On the other hand, neither are new hypotheses unconditionally supported.

Knowledge is augmented through this process. But instead of narrowing the number of possible explanations, the collection of relevant explanations expands. It seems that the research field develops in a *fan-shaped* manner. It widens and spreads out, it covers a larger area, and the amount of knowledge increases. Old hypotheses remain relevant, and at the same time new hypotheses acquire relevance. We know more, but we do not reach a solution.

10.1.2. A complex issue, subject to historical change

Why is no solution is in sight? One reason is that health is a difficult issue. The disease-specific approach, characteristic of medical epidemiology, has

progressed, but it can hardly explain adequately the bulk of common diseases (cf. section 6.5.3). When health is viewed in more sociological and holistic ways, what health is and how it is created become even more complicated.

The issue is even more complex when we are not only addressing individual health, but social inequalities in health, defined as any type of persistent and important differences in aggregated health between social positions in the same social structure(s). Such health inequalities are of a number of types, existing in a number of social structures, in various countries and in different time periods. There is no reason to believe that there will be one single explanation for all these different types of health inequalities. The increasing number of relevant hypotheses may reflect the plurality of the concrete manifestations of social inequalities in health.

Not only the variety of the types of health inequalities, but also their *historical* character, suggest that a search for *the* explanation may be futile. Previously, I have argued that explanations will usually address two themes – the main health determinants believed to produce health inequalities, and the social processes which somehow interact with these health determinants and frame their impact, so that the patterns of health inequalities are generated. But neither the crucial health determinants nor the relevant social processes are historical constants. They change. Health determinants are broad and interrelated groups of factors which vary significantly between social positions. Social and economic change may eliminate the variation in some health determinants and generate new variations in other health determinants. Maybe nutrition and access to health services varied significantly between social classes in mid-20th century. Later, these variations may have diminished in significance, while other health determinants, which formerly acted in more or less the same way across the whole population – health behaviours? psychosocial milieus? – now perhaps display more significant differences between social positions. Similarly as to social processes: social and economic change may modify the prevailing types of social causation processes – structural determination, structural constraint, and the causes for behaviours and lifestyles. New social mobility regimes may appear, modifying how persons with varying health are distributed between social positions.

Over time, this implies that any explanation of how social inequalities in health are generated, whatever its credibility in a particular society at a particular time, may loose its relevance under new conditions. There is therefore not one solution which will have lasting explanatory power. Rather, historical developments will change the critical health determinants and the

relevant social processes, meaning that all explanations must be, so to speak, continuously modernised in order to capture the new circumstances.

10.2. Towards an integrative view?

Currently, what we see within the research field can be understood as attempts to form a more integrative view. It means that ways of explaining health inequalities which *transcend* the dividing lines between the major explanatory approaches are formed. Chapters 4 – 8 described these explanations as more or less "pure" types, and emphasised how they differed from each other. The focus was on oppositions, with social causation on one side and social mobility on the other, and with the physical environments placed as an antithesis to the psychosocial environments. If not earlier, at least during the 1990s the interest in keeping these dividing lines has weakened. Macintyre (1997) contends that in every approach, there are "hard" and "soft" versions. The "hard" versions claim that the preferred explanation is the principal one. The "soft" versions are more modest and claim that the preferred explanation is relevant and contributing, but not necessarily able to account for more than some of the health inequality patterns. In a similar manner, Marmot et al. (1995: 206) suggest that the current situation calls for efforts to form a unifying explanation.

Thus, within the research field, attempts to construct approaches which in some way or another draw on several, even all, of the former explanations described in chapters 4 – 8 can be observed.

The last part of this chapter and of this thesis will discuss two such integrative attempts – the *life course approach*, and the *community approach*. They constitute a new dividing line, not in terms of either social causation or social mobility, or either physical or psychosocial environments, but rather in terms of what types of units they focus on: individuals in a longitudinal perspective, or social wholes and their impact on the health of the members of that whole.

10.2.1. The life course approach

To view people's health as a consequence of their life histories is not new (cf. Kuh and Davey Smith 1993), but this perspective has, by and large, not been at the forefront in health inequalities research. However, as discussed earlier (for instance section 8.3.3), a life course approach to health and health disorders has developed more distinctly in recent decades. It shifts the time

perspective when understanding health, illness and disease from a focus on more immediate antecedents to a longterm view on how the current health of a person has developed. The model may include references to genetic endowment, but will more often point to how biological and social programming in early life (section 8.3.2) constitutes a health potential with enduring consequences for future health. Often, the early formative years, when the organism develops its biological characteristics, is seen as a particularly critical period. Thereafter, health develops as an effect of how exposures to health-damaging and health-enhancing circumstances accumulate over the life course. Thus, although the history of individual health starts with a health potential, it does not end there – health continues to develop because of continuing environmental influences.

This view has been applied in medical epidemiology, with the aim of understanding how risks for specific diseases are influenced by early life influences, followed by later accumulation of exposures. Examples of diseases studied from this perspective are diabetes (McKeigue 1997), respiratory diseases (Strachan 1997), and cardiovascular disease and cancer (Leon and Ben-Shlomo 1997).

But – of special interest for health inequalities research – this life-course view has also been advanced in terms of the *general susceptibility* idea, meaning that accumulated adverse circumstances in the prenatal period, infancy, childhood, adolescence and adulthood may have the effect of a general decline in disease resistance. Alternatively, beneficial circumstances may add to each other as a long series of salutogenic factors. The early years are usually seen as critical. Wadsworth (1997) argues that influences and environments encountered in this period "set parameters" both as regards biology and as regards the ability to resist physical and psychosocial stressors later in life. Wadsworth terms this the person's *health capital* (Wadsworth 1996). It is a start capital, established early in life, but circumstances later on may "adding interest to or depleting from" the initial capital. Thus, the actual health conditions, manifested as specific disorders, may be the result of how the health capital has developed over the life time, due to a long series of adverse or beneficial environments.

This life-course perspective may lead to a more complete understanding of how social inequalities in health are generated. It integrates the health determinants, as it points to how the initial health potential interacts with the later physical and psychosocial environments in creating the current health of a person. It is not only oriented towards questions asked by medical

epidemiology, but may also constitute an explanation of the general pattern of health inequalities via studies of how the early health potential and the varying exposures to physical and psychosocial environments are *socially patterned.*

Various expressions are used to highlight how such a socially patterned exposure to favourable and damaging circumstances may create health inequalities: "the process by which advantages or disadvantages accumulate during the life course" (Blane et al. 1993a: 12), "the unhealthy/unfavourable life career" (Lundberg 1993, Bäckman and Palme 1998), "the co-evolution of health and social achievement" (Vågerö and Illsley 1995), "the accumulation of poor socioeconomic circumstances throughout life" (Hart et al. 1998), and the "transmission of social and biological risk across the life course" (Power et al. 1996b). The common denominator of these expressions is how the current circumstances interact, at every stage in the life course, and also how these current conditions add to or subtract from the previous influences. The actual health risk at any time point is the result of the combined effects of the current situation and the former influences.

This view also opens up for an integration of social causation and social mobility processes. Together, these two types of social processes may form the health inequality patterns. A major idea is of course that health is produced by social causation, as a reflection of the accumulated exposure to more or less adverse or beneficial physical and psychosocial environments. These ideas may be formulated in terms of "chains of risks" (Kuh et al. 1997). There is often a continuity in people's life courses, meaning that a higher risk at some earlier point often will be followed by higher levels of disadvantageous exposures at later points. Thus, adverse material and psychosocial family circumstances, with adverse effects on the health conditions among children, will often be followed by less educational opportunities, less healthy behaviours in adolescence, less advantageous working conditions, less favourable material levels of living in adulthood, and higher frequency of unemployment episodes. Such chains of risks may accumulate to create the marked higher levels of ill health among adults in less advantageous social positions.

The accumulated effects of the environments, in line with the social causation idea, are a basic element of the life course approach to how social inequalities in health are produced. Social mobility processes may however interfere in these life courses, adjusting, aggravating, but perhaps also sometimes breaking, the chains of risks. Depending on how the social mobility regimes function, they may distribute and redistribute the life-time exposures. Processes of indirect or direct health selection will often tend to strengthen the

adverse chains of risk, but also to strengthen the beneficial chains of risks. Social mobility typically takes place in certain life stages: in adolescence in terms of choice of educational careers, and in young adulthood in terms of entrance into the occupational structure, into marriage, and as geographical mobility. Even if the selectivity for health is rather small in such mobility transitions, it may nevertheless add to how the adverse exposures are accumulated in disadvantageous social positions and, vice versa, how the beneficial environments in the more favourable social positions add to each others.

Thus, the life course approach, with its integrated view on how various health determinants interact over the life time, and how social causation and social mobility processes together form the differentiated exposures to health-damaging and health-beneficial circumstances of different social positions, may be a particular fruitful approach for explaining social inequalities.

10.2.2. The community approach

Another current approach within the research field, possibly with the same integrative function, can be termed the *community* approach. A main trait of this approach is that it shifts the focus away from the individual level towards the social systems people are part of.

Wilkinson's and others' suggestions that income inequality has major implications for the level of social cohesion and social trust in society, with further consequences for people's health and the degree of health inequalities, exemplify one line within this trend (cf. section 7.3.4). The general idea these propositions point towards is that it is "something" about the communities people live in – neighbourhoods, cities, regions, countries – that acts on people's health, not only on their average level of health, but also on the differences in health which emerge between various social positions.

For some, the key word is *place*, often meaning the local social context (Popay et al. 1998, see also Curtis and Jones 1998). The focus on place is inspired by various research findings indicating a local determinant of health, which predicts communities' average levels of health and their patterns of health inequalities, significantly in addition to what is explained by individual characteristics (see for instance Karvonen and Rimpela 1997, Curtis and Jones 1998). Others term this social context as *communities*, and refer to studies which "find that community socioeconomic conditions are associated with various measures of health status ... over and above the impact of individual- and family-level socioeconomic position" (Robert and House 2000: 122).

Why this is so, is an intriguing question. Research tries to reflect on what *place* and *communities* mean. They are characterised by their history, their cultural milieus, their modes of social relations, their socioeconomic composition, and their prevailing health-related lay knowledge. Place constitutes the immediate surroundings of people's lives and the circumstances which, in multiple ways, form their everyday existence. Within the context of place, the consequences of physical and psychosocial environments unfold. The prevailing types of social interactions and health behaviours evolve within these local contexts.

The underlying proposition is that within the conditions which together constitute places and communities, the processes which determine health developments – and health inequalities – take place. The focus on place and community does not necessarily imply a very localised notion of how health is created. In the era of globalisation, Popay et al. underline that place is a "location in which macro social structures impact on people's lives" (1998: 619). Macro developments and social forces affect in a multitude of ways the characteristics of places and communities, such as the rate of economic change, the type of workplaces and their variations, and the levels of unemployment (European Science Foundation 2000). Places and communities are meso-level structures, relatively close to people's everyday lives, and macro developments may act through the filter constituted by places and communities.

This approach can be understood as another attempt to formulate an integrated view on the processes which influence health and produce health inequalities. It may be connected to the psychosocial health determinant, as it is in Wilkinson's emphasis on social cohesion and modes of interpersonal relations. It suggests moreover that health-related behaviours, and social variations in these behaviours, are produced by diffusion processes primarily operating at the community level. It may also be connected to the role of physical environments. Immediately, this is seen in factors such as the level of air pollution and the risks entailed in transport and traffic conditions. The physical environments, in terms of material inequalities in housing and material levels of living, may directly influence the creation of health inequalities, but may also act indirectly through their impact on psychosocial milieus. It has also been suggested that local communities and societies are characterised by social capital in quite another tangible way than what Wilkinson proposes. Income levels, as well as income inequality, may be connected to levels of "public investments in education, health care, housing, transportation, (and) environmental quality" (Robert and House 2000: 127, see also Davey Smith 1996). The community level of such assets, as well as community-based

processes which distribute such assets among the population, may have important consequences both for health levels and health inequalities.

Moreover, the community approach may form the basis for more integrated views on the social processes which lie behind the patterns of health inequalities. In a way, it has links to the life course approach, as places and communities constitute the milieus within which life trajectories evolve, and within which the processes of health-related social mobility occur. Also geographical mobility poses interesting questions about how levels of health, and community differences in health, develop. Geographical mobility may be health-related, leading to mobility streams of people with relatively good health into prosperous communities (cf. Boyle et al. 1999), and vice versa, thereby influencing the geographical patterns of health inequalities.

10.3 End remarks

Ideally, a thesis should solve the questions which are posed. This thesis cannot claim to be more than a *contribution* to more clear thinking about what constitutes the research field of social inequalities in health and about the characteristics of explanations and approaches to understand why social inequalities in health emerge. The main question asked in this thesis – how are social inequalities in health generated – is very complex. Moreover, these health inequalities, in spite of their relatively persistent patterns, develop over time. To emphasise the historical character of the subject matter, i.e., how historical change may modify both the health determinants and the relevant social processes, and the time-bounded character of findings and explanations, has been one of the aims. The main question is not easily solved. After decades of intense activity within the research field, Robert and House (2000:115) conclude that "We still do not well and consensually understand, however, why socioeconomic inequalities in health exist and persist, nor what policies are most likely and necessary to reduce these inequalities". Maybe this evaluation is quite apt, both as regards socioeconomic inequalities and also for the other types of health inequalities found in many other social structures.

Social inequalities in health have, during the 19th and 20th centuries and even earlier, been a characteristic part of the health situation in our societies. That this will continue for decades into the 21st century is a probable prognosis. That social change may alter the inequality patterns is not unlikely. How large social change which is necessary to *abolish* them remain in the dark. This thesis has hopefully brought some clarification of the problematic. It nevertheless ends with many question marks.

References

Abbott, P. & Sapsford, R. (1994) Research note: Health and material deprivation in Plymouth: an interim replication, *Sociology of Health & Illness*, 16 (2): 252-259.

Adler, N.E., Boyce, T., Chesney, M.A., Cohen, S., Folkman, S., Kahn, R.I., & Syme, S.L. (1994) Socioeconomic status and health: The challenge of the gradient, *American Psychologist*, 49 (1): 15-24.

Allern, S. (1974) *Klasse og alvorlig sykelighet.* Thesis. Oslo: Instituttet for sosiologi, Universitetet i Oslo.

Aneshensel, C.S. (1992) Social stress: Theory and research, *Annual Review of Sociology*, 18, 15-38.

Andersen, A.S. (1980) Helseforhold, in *Inntektsfordeling og levekår*, SØS nr. 46, Oslo: Statistics Norway.

Annandale, E. (1998) *The sociology of health & medicine. A critical introduction.* Cambridge: Polity Press.

Antonovsky, A. (1979) *Health, stress and coping.* San Francisco: Jossey-Bass Publishers.

Antonovsky. A. (1987) *Unraveling the mystery of health. How people manage stress and stay well.* San Francisco: Jossey-Bass Publishers.

Arber, S. & Lahelma, E. (1993a) Inequalities in women's and men's ill-health: Britain and Finland compared, *Social Science & Medicine*, 37: 1055-1068.

Arber, S. & Lahelma, E. (1993b) Women, paid employment and ill-health in Britain and Finland, *Acta Sociologica*, 36: 121-138.

Arber, S. & Cooper, H. (1999) Gender differences in health in later life: the new paradox? *Social Science & Medicine*, 48 (1): 61-76.

Arntzen, A., Magnus, P. & Bakketeig, L. (1988) Foreldres utdanningslengde relatert til dødelighet rundt fødsel og i første leveår, *Tidsskrift for Den norske lægeforening*, 108:3082-85.

Baird, P.A. (1994) The role of genetics in population health, in Evans, R.G., Barer, M.L. & Marmor, T.R (eds) *Why are some people healthy and others not? The determinants of health of populations.* Berlin/New York: de Gruyter, 133-160.

Bakketeig, L., Cnattingius, S., & Knudsen, L.B. (1993) Socioeconomic differences in fetal and infant mortality in Scandinavia, *Journal of Public Health Policy*, 14: 82-90.

Barker, D.J.P. (ed) (1992) *Fetal and infant origins of adult disease*. London: British Medical Journal.

Barstad, A. (1989) *Legesøkning og sosiokulturelle forhold. Teorier, modeller og empiriske analyser. En litteraturstudie*. INAS-Rapport 89:6, Oslo: INAS.

Bartley, M., & Plewis, I. (1997) Does health-selective mobility account for socioeconomic differences in health? Evidence from England and Wales, 1971 to 1991, *Journal of Health and Social Behavior*, 38: 376-386.

Bartley, M., Popay, J. & Plewis, I. (1992) Domestic conditions, paid employment and women's experience of ill-health, *Sociology of Health & Illness*, 14 (3):313-343.

Bartley, M., Carpenter, L., Dunnel, K. & Fitzpatrick, R. (1996) Measuring inequalities in health: an analysis of mortality patterns using two social classifications, *Sociology of Health & Illness*, 18 (4): 455-474.

Bartley, M., Blane, D. & Montgomery, S. (1997) Socioeconomic determinants of health. Health and the life course: Why safety nets matter, *British Medical Journal*, 314: 1194-1196.

Bartley, M., Sacker, A., Firth, D., & Fitzpatrick, R. (1999) Understanding social variation in cardiovascular risk factors in women and men – The advantage of theoretically based measures, *Social Science & Medicine*, 49 (6): 831-845.

Bates, M.S. & Rankin-Hill, L. (1994) Control, culture and chronic pain, *Social Science & Medicine*, 39: 629-645.

Bauman, Z. (1998) Postmodern adventures of life and death, in Scambler, G. & Higgs, P. (eds) *Modernity, medicine and health. Medical sociology towards 2000.* London - New York: Routledge: 216-231.

Becker, M. H. (1979) Psychosocial aspects of health-related behavior, in Freeman, H.E., Levine, S. & Reeder, L.G. (eds): *Handbook of medical sociology*. New Jersey: Prentice-Hall, Inc.: 253-274.

Bendelow, G. (1993) Pain perceptions, emotions and gender, *Sociology og Health & Illness*, 15 (3): 273-294.

Ben-Shlomo, Y. & Davey Smith, G. (1991) Deprivation in infancy or in adult life: which is more important for mortality risk? *Lancet*, 337 (March 2): 530-534.

Berg, O. (1987) *Medisinens logikk.* Oslo: Universitetsforlaget.

Bjerkedal, T. & Bakketeig, L. (1975) Påliteligheten av intervjuopplysninger om helseforhold, *Tidsskrift for Den norske lægeforening*, 95: 927-931.

Blane, D., Davey Smith, G. & Bartley, M. (1993a) Social selection: what does it contribute to social class differences in health? *Sociology of Health & Illness*, 15 (1):1-15.

Blane, D., Power, C. & Bartley, M. (1993b) The measurement of morbidity in relation to social class, in Abel, T. et al. (eds.) *Medical sociology: research on chronic illness*, Bonn: Informationszentrum Sozialwissenschaften: 65- 84.

Blane, D., White, I. & Morris, J. (1996a) Education, social circumstances, and mortality, in Blane, D., Brunner, E. & Wilkinson, R.G. (eds) *Health and Social Organization. Towards a Health policy for the 21st Century.* London and New York: Routledge: 171-187.

Blane, D., Brunner, E. & Wilkinson, R. (1996b) The evolution of public health policy: an anglocentric view of the last fifty years, in Blane, D., Brunner, E. &

Wilkinson, R.G. (eds) *Health and Social Organization. Towards a Health policy for the 21st Century*. London and New York: Routledge: 1-17.

Blane, D., Bartley, M., & Davey Smith, G. (1997) Disease etiology and materialist explanations of socioeconomic mortality differentials, *European Journal of Public Health*, 7 (4): 385-391.

Blane, D., Montgomery, S.M., & Berney, D.R. (1998) Research note: social class differences in lifetime exposure to environmental hazards, *Sociology of Health & Illness*, 20 (4): 532-536.

Blane, D., Harding, S. & Rosato, M. (1999) Does social mobility affect the size of the socioeconomic mortality differential? Evidence from the Office for National Statistics Longitudinal Study, *Journal of the Royal Statistical Society*, 162 (part 1): 59-70.

Blau, P.M. (1981): Introduction: Diverse views of social structure and their common denominator, in Blau, P.M. & Merton, R.K. (eds) *Continuities in Structural Inquiry*, London and Beverly Hills: Sage Publications: 1-23.

Blaxter, M. (1989) A comparison of measures of inequality in morbidity, in Fox, J. (ed) *Health inequalities in European countries*. Aldershot: Gower: 199-230.

Blaxter, M. (1990) *Health and lifestyles.* London: Routledge.

Blix, E. (1974) Dødelighet og ekteskapelig status i Norge 1969-1970. *Arbeidsnotater 75/45*, Oslo: Statistics Norway.

Bloor, M., Samphier, M. & Prior, L. (1987) Artifact explanations of inequalities in health - an assessment of the evidence, *Sociology of Health & Illness*, 9 (3): 231-264.

Bobak, M. & Marmot, M. (1996) East-West mortality divide and its potential explanations: proposed research agenda, *British Medical Journal*, 312 (17 February): 421-425.

Borgan, J.K. (1995) Socioeconomic trends in differential mortality among middle aged males in Norway 1960-1990, in *Yearbook of population research in Finland*, Papers presented at the Eleventh Nordic Demographic Symposium in Helsinki, Finland 11-13 June, 1995. Helsinki: The Population Research Institute:73-81.

Bottomore, T.B. (1970) *Klass och elit*. Stockholm: Rabén & Sjögren.

Bourdieu, P. (1977) *Outline of a Theory of Practice*. London: Cambridge University Press.

Bourdieu, P. (1984) *Distinction. A social critique of the judgement of taste*. Cambridge, Mass.: Harvard University Press.

Boyle, P. J., Gatrell, A. C. & Duke-Williams, O. (1999) The effect on morbidity of variability in deprivation and population stability in England and Wales – an investigation at small-area level, *Social Science & Medicine*, 49 (6): 791-799.

Brunner, E. (1997) Stress and the biology of inequality, *British Medical Journal*, 314: 1472-1476.

Bunker, J.P, Frazier, H.S. & Mosteller, F. (1994) Improving health: measuring effects of medical care, *Millbank Quarterly*, 72 (2): 225-258.

Bury, M. (1997) *Health and illness in a changing society*. London and New York: Routledge.

Bäckman, O. & Palme, J. (1998) Social background and sickness absence: a study of a Stockholm cohort, *Acta Sociologica*, 41 (4): 349-362.

Carroll, D., Davey Smith, G. & Bennett, P. (1996) Some observations on health and socio-economic status, *Journal of Health Psychology*, 1 (1): 23- 39.

Cavelaars, A. (1998) *Cross-national comparisons of socio-economic differences in health indicators. Thesis*. Rotterdam: Department of Public Health, Erasmus University.

Christensen, L. (1996) Plagsomme pasienter – om sosial konstruksjon i medisinsk allmennpraksis, *Sosiologi i dag*, nr. 3/96: 53-69.

Christophersen, H.O. (1979) *Eilert Sundt – En dikter i kjensgjerninger*. Oslo: Gyldendal Norsk Forlag – Fakkel-serien.

Claussen, B. (1999) Health and reemployment in a 5-year follow-up of long-term unemployed, *Scandinavian Journal of Public Health*, 27 (2): 94-100.

Cockerham, W.C., Rütten, A. & Abel, T. (1997) Conceptualizing contemporary health lifestyles: moving beyond Weber, *Sociological Quarterly*, 38 (2): 321-342.

Conrad, P. (1999) A mirage of genes, *Sociology of Health & Illness*, 21 (2): 228-241.

Conrad, P. & Gabe, J. (1999) Introduction: Sociological perspectives on the new genetics: an overview, *Sociology of Health & Illness*, 21 (5): 505-516.

Crawford, R. (1977) You are dangerous to your health - the ideology and politics of victim blaming, *International Journal of Health Services*, 7(4): 663-679.

Crompton, R. (1996) The fragmentation of class analysis, *British Journal of Sociology*, 47 (1): 56-67.

Curtis, S. & Jones, I.R. (1998) Is there a place for geography in the analysis of health inequality? *Sociology of Health & Illness*, 20 (5): 645-672.

Dahl, E. (1991) Inequality in health and the class position of women - the Norwegian experience, *Sociology of Health & Illness*, 13 (4): 492-505.

Dahl, E. (1993a) High mortality in lower salaried Norwegian men: the healthy worker effect? *Journal of Epidemiology and Community Health*, 47:153-157.

Dahl, E. (1993b) Social inequality in health – the role of the healthy worker effect, *Social Science & Medicine*, 36 (8):1077-1086.

Dahl, E. (1994a) Social inequalities in ill-health – the significance of occupational status, education and income. Results from a Norwegian survey, *Sociology of Health & Illness*, 16 (5): 644-667.

Dahl, E. (1994b) *Sosial ulikhet i helse: Artefakter eller seleksjon? Dissertation*. FAFO-Rapport 170. Oslo: FAFO.

Dahl, E. (1996) Sosial ulikhet i helse – en diskusjon av seleksjonsforklaringen, in Lian, O. S. (ed) *Helsetjenesten i samfunnsvitenskapens lys*. Oslo: Tano Aschehoug: 47-59.

Dahl, E. & Kjærsgaard, P. (1993) Trends in socioeconomic mortality differentials in post-war Norway - evidence and interpretations, *Sociology of Health & Illness*, 15(5): 447-471.

Dahl, E. & Birkelund, G. (1997) Health inequalities among the elderly in a socialdemocratic welfare state, *Social Science & Medicine*, 44 (6): 871-881.

Dahl, E. & Birkelund, G. (1999) Sysselsetting, klasse og helse 1980-1995, *Tidsskrift for samfunnsforskning*, 40 (1): 3-32.

Dahl, E. & Elstad, J.I. (1999) Recent changes in social structure and health inequalities in Norway, *Draft.* Oslo: FAFO-NOVA.

Davey Smith, G. (1996) Income inequality and mortality: why are they related? *British Medical Journal*, 312 (20 April): 987-988.

Davey Smith, G. (1997) Socioeconomic differentials, in Kuh, D. & Ben-Shlomo, Y. (eds) *A Life Course Approach to Chronic Disease Epidemiology*, Oxford New York Tokyo: Oxford University Press: 242-273.

Davey Smith, G. & Phillips, A. N. (1992) Confounding in epidemiological studies: why "independent" effects may not be all they seem, *British Medical Journal*, 305 (26 September): 757-759.

Davey Smith, G., Hart, C., Blane, D., Gillis, C., & Hawthorne, V. (1997) Lifetime socioeconomic position and mortality: prospective observational study, *British Medical Journal*, 314 (22 February): 547-552.

Davey Smith, G., Hart, C., Blane, D. & Hole, D. (1998) Adverse socioeconomic conditions in childhood and cause specific adult mortality: prospective observational study, *British Medical Journal*, 316 (30 May): 1631-1635.

Davison, C., Fankel, S. & Davey Smith, G. (1992) The limits of lifestyle: re-assessing 'fatalism' in the popular culture of illness prevention *Social Science & Medicine*, 34 (6): 675-685.

DHSS (1980) *Inequalities in Health: Report of a Working Group Chaired by Sir Douglas Black.* London: Department of Health and Social Security.

Dubos, R. (1995[1979]) The mirage of health, in Davey, B., Gray, A. & Seale C. (eds.) *Health and disease. A reader.* Buckingham - Philadelphia: Open University Press: 4-10.

Durkheim, E. (1964[1893]) *The Division of Labor in Society.* New York: Free Press.

Ecob, R. & Davey Smith, G. (1999) Income and health – what is the nature of the relationship? *Social Science & Medicine*, 48 (5): 693-705.

Elstad, J.I. (1981) Kroniske lidelser og sosial klasse. *Arbeidsnotater*, nr. 170, Oslo: Instituttet for sosiologi, Universitetet i Oslo.

Elstad, J.I. (1995a) Employment status and women's health – exploring the dynamics, *Acta Sociologica*, 38 (3): 231-249.

Elstad, J.I. (1995b) Sosial klasse i helsestatistikk og forskning, *Tidsskrift for Den norske lægeforening*, 115 (4): 502-504.

Elstad, J.I. (1996a) Inequalities in health related to women's marital, parental, and employment status – a comparison between the early 70s and the late 80s, Norway, *Social Science & Medicine*, 42 (1): 75-89.

Elstad, J.I. (1996b) How large are the differences – really? Self-reported long-standing illness among working class and middle class men", *Sociology of Health & Illness*, 18 (4): 475-498.

Elstad, J.I. (1998a) *Funksjonshemmede og psykisk helse*. NOVA-Rapport 3/98. Oslo: NOVA.

Elstad, J.I. (1998b) The psycho-social perspective on social inequalities in health, *Sociology of Health & Illness*, 20 (5): 598-618.

Elstad, J.I. (1999a) *Kjønnsforskjeller i helse*. NOVA-Rapport 13/99, Oslo: NOVA.

Elstad, J.I. (1999b) *From poverty to prosperity*. Unpublished draft. Oslo: NOVA.

Elstad, J.I. (2001) Health-related mobility, health inequalities, and gradient constraint: Discussion and results from a Norwegian study, *European Journal of Public Health*,. forthcoming.

Engels, Frederick (1969[1845]) *The conditions of the working class in England.* Frogmore, St Albans, Herts: Panther Books Ltd.

Erikson, R. & Goldthorpe, J. (1992) *The constant flux. A study of class mobility in industrial societies*. Oxford: Clarendon Press.

Esping-Andersen, G. (ed) (1993) *Changing classes. Stratification and mobility in post-industrial societies*. London: SAGE.

European Science Foundation (2000) *Social Variations in Health Expectancy in Europe*, www.uni-duesseldorf/health.

Evang, K. (1974) *Helse og samfunn. Sosialmedisinsk allmennkunnskap*. Oslo: Gyldendal Norsk Forlag.

Evans, R. G. (1994) Introduction, in Evans, R.G., Barer, M.L. & Marmor, T.R. (eds) *Why are some people healthy and others not? The determinants of health of populations.* Berlin/New York: de Gruyter: 3-26.

Evans, R. G., and Stoddart, G. L. (1994) Producing health, consuming health care, in Evans, R.G., Barer, M.L. & Marmor, T.R. (eds) *Why are some people healthy and others not? The determinants of health of populations.* Berlin/New York: de Gruyter: 27-65.

Evans, R.G., Barer, M.L. & Marmor, T.R. (1994a) Preface, in Evans, R.G., Barer, M.L. & Marmor, T.R. (eds) *Why are some people healthy and others not? The determinants of health of populations.* Berlin/New York: de Gruyter: ix-xix.

Evans, R.G., Hodge, M. & Pless, I.B. (1994b) If not genetics, then what? Biological pathways and population health, in Evans, R.G., Barer, M.L. & Marmor, T.R. (eds) *Why are some people healthy and others not? The determinants of health of populations.* Berlin/New York: de Gruyter: 161-188.

Finnvold, J.E. & Nordhagen, R. (1996) Helseulikhet i barns første leveår, *Samfunnsspeilet* 2/96: 17-24.

Fitzpatrick, R. (1984) Lay concepts of illness, in Fitzpatrick, R. et al. (eds) *The Experience of Illness*, London: Tavistock Publications.

Fitzpatrick, R.M. (1997) Society and changing patterns of disease, in Scambler, G. (ed) *Sociology as applied to medicine. 4th edition*, London: W.B.Saunders Company Ltd.: 3-17.

Fitzpatrick, R. & Scambler, G. (1984) Social class, ethnicity, and illness, in Fitzpatrick, R. et al. (eds) *The Experience of Illness*, London: Tavistock Publications: 54-84.

Fonagy, P. (1996) Patterns of attachment, interpersonal relationships and health, in Blane, D., Brunner, E. & Wilkinson, R.G. (eds) *Health and Social Organization. Towards a Health policy for the 21st Century*. London and New York: Routledge: 125-151.

Forsdahl, A. (1977) Are poor living conditions in childhood and adolescence an important risk factor for arteriosclerotic heart disease? *British Journal of Preventive and Social Medicine*, 31: 91-95.

Fox, J. (ed) (1989) *Health inequalities in European countries*. Aldershot: Gower.

Fox, J. (1990) Social class, mental illness, and social mobility: The social selection-drift hypothesis for serious mental illness, *Journal of Health and Social Behavior*, 31 (December): 344-353.

Freidson, E. (1970) *Profession of medicine*, New York: Dodd, Mead.

Freund, P. (1982) *The civilized body. Social domination, control, and health.* Philadelphia: Temple University Press.

Gerhardt, U. (1989) *Ideas about illness. An intellectual and political history of medical sociology*. London: Macmillan.

Gerth, H.H. & Mills, C.W. (1970) *From Max Weber. Essays in sociology*, London: Routledge & Kegan Paul Ltd.

Giddens, A. (1973) *The class structure of advanced societies*. London: Hutchinson.

Giddens, A. (1984) *The constitution of society*. Cambridge: Polity Press.

Gissler, M., Rahkonen, O., Jarvelin, M.R. & Hemminki, E. (1998) Social class differences in health until the age of 7 years among the Finnish 1987 birth cohort, Abstract, *Social Science & Medicine*, 46 (12): 1543-1552.

Glendinning, A., Shucksmith, J. & Hendry, L. (1994) Social class and adolescent smoking behaviour, *Social Science & Medicine*, 38 (10): 1449-1460.

Goldberg, E.M. & Morrison, S.L. (1963) Schizophrenia and social class, *British Journal of Psychiatry*, 109: 785-802.

Gravelle, H. (1998) How much of the relation between population mortality and unequal distribution of income is a statistical artefact, *British Medical Journal*, 316: 382-385.

Grøtvedt, L. & Belsby, L. (1995) *Barns helse*. Rapporter 95/11. Oslo – Kongsvinger: Statistics Norway.

Grue, L. (1995) *Funksjonshemmet - klassifisering og definisjoner*. INAS-Notat 95:1. Oslo: INAS.

Habermas. J. (1969) Erkjennelse og interesse, in Habermas, J. *Vitenskap som ideologi*. Oslo: Gyldendal Norsk Forlag: 3-17.

Halldorsson, M., Cavelaars, A., Kunst, A. & Mackenbach, J.P. (1999) Socioeconomic differences in health and well-being of children and adolescents in Iceland, *Scandinavian Journal of Public Health*, 27 (1): 43-47.

Hansen, F.H. (1979) Helsesektoren i velferdsstaten: kjempevekst og fordelingskrise, *Tidsskrift for samfunnsforskning*, 20 (3): 219-240.

Harkey, J., Miles, D.L. & Rushing, W.A. (1976) The relation between social class and functional status: a new look at the drift hypothesis, *Journal of Health and Social Behavior*, 17 (September): 194-204.

Hart, C., Davey Smith, G., & Blane, D. (1998) Inequalities in mortality by social class measured at 3 stages of the lifecourse, *American Journal of Public Health*, 88 (3): 471-474.

Hellevik, O. (1997) Class inequality and egalitarian reform, *Acta Sociologica*, 40 (4): 377-397.

Hellevik, O. (2000) A less biased allocation mechanism, *Acta Sociologica*, 43 (1): 81-83.

Hemström, Ö. (1996) Is marriage dissolution linked to differences in mortality risks for men and women? *Journal of Marriage and the Family*, 58 (May '96): 366-378.

Hertzman, C., Frank, J., & Evans, R.G. (1994) Heterogeneities in health status, in Evans, R.G., Barer, M.L. & Marmor, T.R. (eds) *Why are some people healthy and others not? The determinants of health of populations.* Berlin/New York: de Gruyter: 67-92.

Higgs, P. & Scambler, G. (1998) Explaining health inequalities. How useful are concepts of social class? In Scambler, G. & Higgs, P. (eds) *Modernity, medicine and health. Medical sociology towards 2000*. London & New York: Routledge: 82-99.

Himsworth, H. (1984) Epidemiology, genetics and sociology, *Journal of Biosocial Science*, 16: 159-176.

Hodne, F. (1986) Medisin og miljø: nye synspunkter, in Larsen, Ø., Berg, O. & Hodne, F. (eds) *Legene og samfunnet*, Oslo: Seksjon for medisinsk historie og Den norske lægeforening: 76-92.

Holme, I., Helgeland, A, Hjermann, I, Leren, P., & Lund-Larsen, P.G. (1980) Four-year mortality by some socioeconomic indicators: the Oslo study, *Journal of Epidemiology and Community Health*, 34:48-52.

Holstein, B.E., Iversen, L. & Kristensen, T.S. (1988) *Medicinsk sociologi*. København – Århus – Odense: Foreningen af danske lægestuderendes forlag.

Hovland, A. (1999) *Moderne urfolk – lokal og etnisk tilhørighet blant samisk ungdom*. NOVA-Rapport 11/99, Oslo: NOVA.

Idler, E.L. & Benyamini, Y. (1997) Self-rated health and mortality – a review of 27 community studies, *Journal of Health and Social Behavior*, 38 (1):21-37.

Illich, I. (1975) *Den farliga sjukvården*. Stockholm: Aldusserien, 439.

Illsley, R. (1955) Social class selection and class differences in relation to stillbirths and infant deaths, *British Medical Journal*, 2:1520-1526.

Illsley, R. (1986) Occupational class, selection and the production of inequalities in health, *Quarterly Journal of Social Affairs*, 2 (2): 151-165.

Illsley, R. (1987) Occupational class, selection and inequalities in health: rejoinder to Richard Wilkinson's reply, *Quarterly Journal of Social Affairs*, 3 (3): 213-223.

Illsley, R. & Svensson, P.G. (1984) *The health burden of social inequities. Country profiles from 20 member states of the European Region of the World Health Organization*. Copenhagen: World Health Organization, Regional office for Europe.

Joffe, M. (1989) Social inequalities in low birth weight: timing of effects and selective mobility, *Social Science & Medicine*, 28 (6): 613-619.

Jozan, P. (1989) A compilation of some aspects of area mortality differentials in some European countries, in Fox, J. (ed) *Health inequalities in European countries*. Aldershot: Gower: 173-196.

Judge, K. (1995) Income distribution and life expectancy: a critical appraisal, *British Medical Journal*, 311 (11 November): 1282-1285.

Judge, K., Mulligan, J. & Benzeval, M. (1998) Income inequality and population health, *Social Science & Medicine*, 46 (4-5): 567-569.

Kahn, H.S., Tatham, L.M., Pamuk, E.R., & Heath, C.W. (1998) Are geographic regions with high income inequality associated with risk of abdominal weight-gain? *Social Science & Medicine*, 47 (1): 1-6.

Kaplan. G. A., Pamuk, E. R., Lynch, J. W., Cohen, R. D., & Balfour, J. L. (1996) Inequality in income and mortality in the United States: analysis of mortality and potential pathways, *British Medical Journal*, 312 (20 April): 999-1003.

Karasek, R. & Theorell, T. (1990) *Healthy work: stress, productivity, and the reconstruction of working life*. New York: Basic Books.

Karvonen, S. & Rimpela, A.H. (1997) Urban small-area variation in adolescents health behaviour, *Social Science & Medicine*, 45 (7): 1089-1098.

Karvonen, S., Rimpela, A.H., & Rimpela, M.I. (1999) Social mobility and health-related behaviours in young people, *Journal of Epidemiology and Community Health*, 53 (4): 211-217.

Katz-Gerro, T. & Shavit, Y. (1998) The stratification of leisure and taste: Classes and lifestyles in Israel, *European Sociological Review*, 14 (4): 369-386.

Kawachi, I. & Kennedy, B. P. (1997) Health and social cohesion: why care about income inequality, *British Medical Journal*, 314 (5 April): 1037-1040.

Kawachi, I., Kennedy, B. P., Lochner, K., & Prothrow-Stith, D. (1997) Social capital, income inequality, and mortality, *American Journal of Public Health*, 87 (9): 1491-1498.

Kelly, S., Hertzman, C. & Daniels, M. (1997) Searching for the biological pathways between stress and health, *Annual Review of Public Health*, 18: 437-462.

Kendall, P.L. & Reader, G.G. (1972) Contributions of sociology to medicine, in Freeman, H.E., Levine, S. & Reeder, L.G. (eds): *Handbook of medicial sociology*. New Jersey: Prentice-Hall, Inc.:1-22.

Kennedy, B.P., Kawachi, I. & Prothrow-Stith, D. (1996) Income distribution and mortality: cross sectional ecological study of the Robin Hood index in the United States, *British Medical Journal*, 312(20 April): 1006-1007.

Killén, K. (1999) Tidlig mor-barn samspill og tilknytning, *Tidsskrift for Norsk Psykologforening*, 36: 938-948.

Kjøller, M., Rasmussen, N.K., Keiding, L., Petersen, H.C. & Nielsen, G.A. (1995) *Sundhed og sykelighed i Danmark 1994 – og udviklingen siden 1987*, København: Dansk Institut for Klinisk Epidemiologi.

Kock, W. (1955) *Medicinhistoriens grunddrag*. Stockholm: Natur och Kultur.

Kohler, L. (ed) *Barn och barnfamiljer i Norden. En studie av välfärd, hälsa och livskvalitet*. NHV-Rapport 1, Göteborg: Nordiska Hälsovårdhögskolan.

Koopman, C., Eisenthal, S. & Stoeckle, J.D. (1984) Ethnicity in the reported pain, emotional distress and requests of medical outpatients, *Social Science & Medicine*, 18 (6): 487-490.

Krieger, N., Chen, J.T. & Selby, J.V. (1999) Comparing individual based and household based classifications of women's health, *Journal of Epidemiology and Community Health,* 53 (10): 612-623.

Kringlen, E. (1986) *Medisin og samfunn*. Oslo Bergen Stavanger Tromsø: Universitetsforlaget.

Kristofersen, L. B. (1979) *Yrke og dødelighet. Tilleggsberegninger til tidligere analyse*. Rapporter 79/19, Oslo: Statistics Norway.

Kruger, O., Aase, A. & Westin, S. (1995) Ischemic heart disease mortality among men in Norway - reversal of urban-rural difference between 1966 and 1989, *Journal of Epidemiology and Community Health,* 49 (3): 271-276.

Kuh, D. & Davey Smith, G. (1993) When is mortality risk determined? Historical insights into a current debate, *Social History of Medicine*, 6:101-123.

Kuh, D. & Davey Smith, G. (1997) The life course and adult chronic disease: an historical perspective with particular reference to coronary heart disease, in Kuh, D. & Ben-Shlomo, Y. (eds) *A Life Course Approach to Chronic Disease Epidemiology*, Oxford New York Tokyo: Oxford University Press: 15-41.

Kuh, D., Power, C., Blane, D. & Bartley, M. (1997) Social pathways between childhood and aduilt health, in Kuh, D. & Ben-Shlomo, Y. (eds) *A Life Course*

Approach to Chronic Disease Epidemiology, Oxford New York Tokyo: Oxford University Press: 169-198.

Kunst, A. (1997) *Cross-national comparisons of socio-economic differences in mortality. Thesis*. Rotterdam: Department of Public Health, Erasmus University.

Kunst, A., Groenhof, F., Borgan, J.K., Costa, G., Desplanques, G., Faggiano, F., Hemström, O., Martikainen, P., Vågerø, D., Valkonen, T. & Mackenbach, J.P. (1998) Socioeconomic inequalities in mortality – methodological problems illustrated with 3 examples from Europe, Abstract, *Revue d'Epidemiologie et de Santé Public*, 46 (6): 467-479.

Lahelma, E. (1997) *Social variations in health: Nordic comparisons and changes over time*. Research Proposal, Helsinki: Department of Public Health, University of Helsinki.

Lahelma, E., Martikainen, P., Rahkonen, O. & Silventoinen, K. (1999) Gender differences in illhealth in Finland: patterns, magnitude and change, *Social Science & Medicine*, 48 (1): 7-19.

Lahelma, E., Arber, S., Rahkonen, O. & Silventoinen, K. (2000) Widening or narrowing inequalities in health – comparing Britain and Finland from the 1980s to the 1990s, *Sociology of Health & Illness*, 22 (1): 110-136.

Landsbergis, P.A., Schnall, P.L., Deitz, D.K., Warren, K., Pickering, T.G. & Schwartz, J.E. (1998) Job strain and health behaviors. Results of a prospective study, *American Journal of Health Promotion*, 12 (4): 237-245.

Leach, E.R. (1981) British social anthropology and Lévi-Straussian structuralism, in Blau, P.M. & Merton, R.K. (eds) *Continuities in Structural Inquiry*, London and Beverly Hills: Sage Publications: 27-49.

Leon, D. & Ben-Shlomo, Y. (1997) Preadult influences on cardiovascular disease and cancer, in Kuh, D. & Ben-Shlomo, Y. (eds) *A Life Course Approach to Chronic Disease Epidemiology*, Oxford New York Tokyo: Oxford University Press: 45-77.

Li, J.H. & Singelmann, J. (1998) Gender differences in class mobility: a comparative study of the United States, Sweden, and West Germany, *Acta Sociologica*, 41 (4): 315-333.

Link, B.G. & Phelan, J. (1995) Social conditions as fundamental causes of disease, *Journal of Health and Social Behavior*, Extra Issue: 80-94.

Link, B.G., Northridge, M.E., Phelan, J.C., & Ganz, M.L. (1998) Social epidemiology and the fundamental cause concept: On the structuring of effective cancer screens by socioeconomic status, *Millbank Quarterly*, 76 (3): 375-402.

Locker, D. (1997) Social causes of disease, in Scambler, G. (ed) *Sociology as applied to medicine*. 4th edition, London: W.B.Saunders Company Ltd.: 18-31.

Lundberg, O. (1991) Childhood living conditions, health status, and social mobility: a contribution to the health selection debate, *European Sociological Review*, 7 (2): 149-162.

Lundberg, O. (1993) The impact of childhood living conditions on illness and mortality in adulthood, *Social Science & Medicine*, 36 (8): 1047-1052.

Lundberg, O. (1997) Childhood conditions, sense of coherence, social class and adult ill health: exploring their theoretical and empirical relations, *Social Science & Medicine*, 44 (6): 821-831.

Lynch, J.W., Kaplan, G.A., Cohen, R.D., Kauhanen, J., Wilson, T.W., Smith, N.L. & Salonen, J.T. (1994) Childhood and adult socioeconomic status as predictors of mortality in Finland, *Lancet*, 343 (February 26): 524-527.

Lynch, J.W., Kaplan, G.A. & Salonen, J.T. (1997) Why do poor people behave poorly? Variation in adult health behaviours and psychosocial characteristics by stages of the socioeconomic lifecourse, *Social Science & Medicine*, 44 (6): 809-819.

Lynch, J.W. & Kaplan, G.A. (1997) Wither studies on the socioeconomic foundations of population health? *American Journal of Public Health*, 87: 1409-1411.

Mackenbach, J.P. & Kunst, A.E. (1997) Measuring the magnitude of socio-economic inequalities in health: an overview of available measures illustrated with two examples from Europe, *Social Science & Medicine*, 44 (6): 757-771.

Macintyre, S. (1997) The Black Report and beyond: What are the issues? *Social Science & Medicine*, 44 (6): 723-745.

Macintyre, S. & West, P. (1991) Social, developmental and health correlates of 'attractiveness' in adolescence, *Sociology of Health & Illness*, 13 (2):149-167.

Macintyre, S., Ford, G., & Hunt, K. (1999) Do women 'over-report' morbidity? Men's and women's responses to structured prompting on a standard question on long standing illness, *Social Science & Medicine*, 48 (1): 89-98.

Manor, O., Matthews, S. & Power, C. (1997) Comparing measures of health inequality, *Social Science & Medicine*, 45 (5):761-771.

Marmot, M. (1994) Social differences in health within and between populations, *Dædalus*, 123 (4) (Fall): 197-216.

Marmot, M. (1996) The social pattern of health and disease, in Blane, D., Brunner, E. & Wilkinson, R.G. (eds) *Health and Social Organization. Towards a Health policy for the 21st Century*. London and New York: Routledge: 42-67.

Marmot, M. & Davey Smith, G. (1989) Why are the Japanese living longer, *British Medical Journal*, 299:1547-1551.

Marmot, M., Adelstein, A.M., Robinson, N., & Rose, G.A. (1978) Changing social class distribution of heart disease, *British Medical Journal*, ii: 1109-1112.

Marmot, M.G., Bobak, M. & Davey Smith, G. (1995) Explanations for social inequalities in health, in Amick III B.C., Levine, S., Tarlov, A.R., & Chapman Walsh, D. (eds) *Society and Health*. London: Oxford University Press: 172-210.

Marmot, M., Ryff, C.D., Bumpass, L.L., Shipley, M. & Marks, N.F. (1997) Social inequalities in health - next questions and converging evidence, *Social Science & Medicine*, 44 (6): 901-910.

Marshall, G. & Swift, A. (1999) On the meaning and measurement of inequality, *Acta Sociologica*, 42 (3): 241-250.

Martin, P. (1999) Genes as drugs: the social shaping of gene therapy and the reconstruction of genetic disease, *Sociology of Health & Illness*, 21(5): 517-538.

Mastekaasa, A. (1993a) *The relationship between marital status and subjective well-being: consistency, variations, and causal explanations*. Dissertation, ISO-Rapport nr. 38 1993, Oslo: Instituttet for sosiologi, Universitetet i Oslo.

Mastekaasa, A. (1993b) Marital dissolution and subjective distress: Panel evidence, in Mastekaasa, A. *The relationship between marital status and subjective well-being: consistency, variations, and causal explanations*. Dissertation, ISO-Rapport nr. 38 1993, Oslo: Instituttet for sosiologi, Universitetet i Oslo: Study H.

McKeigue, P. (1997) Diabetes and insulin action, in Kuh, D. & Ben-Shlomo, Y. (eds) *A Life Course Approach to Chronic Disease Epidemiology*, Oxford New York Tokyo: Oxford University Press: 78-100.

McKeown, T. (1979) *The role of medicine*. Oxford: Basil Blackwell.

Mechanic, D. (1978) *Medical sociology. A comprehensive text*. 2nd Edition. New York London: The Free Press.

Miech, R.A., Caspi, A., Mofitt, T.E., Wright, B.R.E. & Silva, P.A. (1999) Low socioeconomic status and mental disorders: a longitudinal study of selection and causation during young adulthood, *American Journal of Sociology*, 104 (4): 1096-1131.

Mheen, H. van de, Stronks, K., & Mackenbach, J.P. (1998) A lifecourse perspective on socio-economic inequalities in health: the influence of childhood socio-economic conditions and selection processes, *Sociology of Health & Illness*, 20 (5): 754-777.

Mheen, H. van de, Stronks, K., Schrijvers, C.T.M., & Mackenbach, J.P. (1999) The influence of adult ill health on occupational class mobility and mobility out of and into employment in the Netherlands, *Social Science & Medicine*, 49 (4): 509-518.

Moum, T. (1998) Mode of administration and interviewer effects in self-reported symptoms of anxiety and depression, *Social Indicators Research*, 45: 279-318.

Muntaner, C. & Lynch, J. (1999) Income inequality, social cohesion, and class relations: a critique of Wilkinson's neo-Durkheimian research program, *International Journal of Health Services*, 29 (1): 59-81.

Mønnesland, J., Brunborg, H. & Selmer, R. (1982) *Inngåelse og oppløsning av ekteskap etter alder og varighet*. Rapporter 82/2, Oslo: Statistics Norway.

Nazroo, J. (1998) Genetic, cultural or socioeconomic vulnerability? Explaining ethnic inequalities in health, *Sociology of Health & Illness*, 20 (5): 710-730.

Navarro, V. (1989) Race *or* class, or race *and* class, *International Journal of Health Services*, 19 (2): 311-314.

Nettleton, S. (1995) *The sociology of health & illness*. London: Polity Press.

Nordhagen, R. & Bakketeig, L.S. (1994) I begynnelsen var barnet... Avgjøres helse og livsløp allerede i fosterliv og tidlig barndom? *Tidsskrift for Den norske lægeforening*, 114 (30): 3570-3575.

Næss, S. & Bergwitz, Ø. (1991) Ekteskapet – best for menn? *Tidsskrift for samfunnsforskning*, 32 (4): 313-329.

Osler, M. (1993) Social class and health behavior in Danish adults – a longitudinal study, *Public Health*, 107 (4): 251-260.

Pakulski, J. & Waters, M. (1996) *The Death of Class*. London: SAGE.

Pearlin, L. I. (1989) The sociological study of stress, *Journal of Health and Social Behavior*, 30 (September): 241-256.

Peck, M.N. (1994) The importance of childhood socio-economic group for adult health, *Social Science & Medicine*, 39 (4): 553-562.

Peck, M. N. & Vågerö, D.H. (1989) Adult body height, self perceived health and mortality in the Swedish population, *Journal of Epidemioogy and CommunityHealth*, 43: 380-384.

Pill, R. & Stott, N.C. (1982) Concepts of illness causation and responsibility: some preliminary data from a sample of working class mothers, *Social Science & Medicine*, 16 (1): 43-52.

Popay, J., Williams, G., Thomas, C. & Gatrell, A. (1998) Theorising inequalities in health: the place of lay knowledge, *Sociology of Health & Illness*, 20 (5): 619-645.

Power, C., Matthews, S., & Manor, O. (1996a) Inequalities in self-rated health in the 1956 birth cohort: lifetime social circumstances or social mobility? *British Medical Journal*, 313: 449-453.

Power, C., Bartley, M., Davey Smith, G., & Blane, D. (1996b) Transmission of social and biological risk across the life course, in Blane, D., Brunner, E. & Wilkinson, R.G. (eds) *Health and Social Organization. Towards a Health policy for the 21st Century.* London and New York: Routledge: 188-203.

Prandy, K. (1999) Class, stratification and inequalities in health. A comparison of the Registrar General's Social Classes and the Cambridge Scale, *Sociology of Health & Illness*, 21 (4): 466-484.

Rahkonen, O., Lahelma, E., & Huuka, M. (1997a) Past of present? Childhood living conditions and current socioeconomic status as determinants of adult health, *Social Science & Medicine*, 44 (3): 327-336.

Rahkonen, O., Arber, S. & Lahelma, E. (1997b) Health-related social mobility: a comparison of currently employed men and women in Britain and Finland, *Scandinavian Journal of Social Medicine*, 25 (2): 83-92.

Rahkonen, O., Arber, S., Lahelma, E., Martikainen, P., & Silventoinen, K. (2000) Understanding income inequalities in health among men and women in Britain and Finland, *International Journal of Health Services*, 30 (1): 27-48.

Ramsøy, N.R. (1977) *Sosial mobilitet i Norge*. Oslo: INAS-Tiden.

Renaud, M. (1994) The future: Hygeia versus Panakeia, in Evans, R.G., Barer, M.L. & Marmor, T.R (eds) *Why are some people healthy and others not? The determinants of health of populations.* Berlin/New York: de Gruyter: 317-334.

Ringdal, K. (1994) Intergenerational class mobility in post-war Norway: a weakening of vertical barriers? *European Sociological Review*, 10 (3): 273-288.

Ringen, S. (1976) *Om helseforhold.* Arbeidsnotat 83, Bergen: Levekårsundersøkelsen 1973.

Ritzer, G. (1975) *Fundamentale perspektiver i sociologien.* Odense: Fremad.

Ritzer, G. (1992) *Contemporary Sociological Theory. Third Edition.* New York: McGraw-Hill, Inc.

Robert, S.A. & House, J. (2000) Socioeconomic inequalities in health: integrating individual-, community-, and societal-level theory and research, in Albrecht, G.L., Fitzpatrick, R. & Scrimshaw, S.C. (eds) *Handbook of social studies in health and medicine*, London Thousand Oaks New Delhi: SAGE Publications: 115-135.

Rodgers, G.B. (1979) Income and inequality as determinants of mortality: an international cross-sectional analysis, *Population Studies*, 33: 343-351.

Rognerud, M. & Stensvold, I. (eds) (1998) *Oslohelsa. Utredningen om helse, miljø og sosial ulikhet i bydelene.* Oslo: Ullevål sykehus, Klinikk for forebyggende medisin.

Rose, G. (1981) Strategy of prevention: lessons from cardiovascular disease, *British Medical Journal*, 282: 1847-1851.

Rose, G. (1985) Sick individuals and sick populations, *International Journal of Epidemiology*, 14: 32-38.

Samson, C. (1999) Biomedicine and the body, in Samson, C. (ed) *Health Studies. A Critical and Cross-Cultural Reader*, Oxford: Blackwell Publishers: 3-21.

Schalick, L.M., Hadden, W.C., Pamuk, E., Navarro, V. & Pappas, G. (2000) The widening gap in death rates among income groups in the United States from 1967 to 1986, *International Journal of Health Services*, 30 (1): 13-26.

Sihvonen, A.P., Kunst, A.E., Lahelma, E., Valkonen, T. & Mackenbach, J.P. (1998) Socioeconomic inequalities in health expectancy in Finland and Norway in the late 1980s, *Social Science & Medicine*, 47 (3): 303-315.

Silventoinen, K., Lahelma, E. & Rahkonen, O. (1998) Social background, adult body-height and health, *Paper*, Helsinki: University of Helsinki.

Smaje, C. (1995) The ethnic patterning of health: new directions for theory and research, *Sociology of Health & Illness*, 18 (2): 251-269.

Soobader, M.J. & Leclere, F.B. (1999) Aggregation and the measurement of income inequality – effects on morbidity, Abstract, *Social Science & Medicine*, 48 (6): 733-744.

Sorokin, P. (1964[1927]) *Social and cultural mobility.* New York: The Free Press of Glencoe.

Sosial- og helsedepartementet (1996) *Folkehelserapporten* [Population Health Report]. Oslo: Sosial- og helsedepartementet.

Statistics Norway (1969) Dødelighet i fylkene. Regional dødelighet. *NOS A 298.* Oslo: Statistics Norway.

Statistics Norway (1993) Standarder for norsk statistikk. Klassifikasjon av sykdommer, skader og dødsårsaker. Systematisk del. Norsk utgave av ICD-9. *NOS C 107*. Oslo-Kongsvinger: Statistics Norway.

Statistics Norway (1998a) *Statistisk Årbok 1998*. Oslo-Kongsvinger: Statistics Norway.

Statistics Norway (1998b) *Dødsårsaker 1995*. Oslo-Kongsvinger: Statistics Norway.

Stern, J. (1983) Social mobility and the interpretation of social class mortality differentials, *Journal of Social Policy*, 12: 27-49.

Strachan, D.P (1997) Respiratory and allergic diseases, in Kuh, D. & Ben-Shlomo, Y. (eds) *A Life Course Approach to Chronic Disease Epidemiology*, Oxford New York Tokyo: Oxford University Press: 101-120.

Svensson, R. (1993) *Samhälle, medicin, vård. En introduktion till medicinsk sociologi*. Lund: Studentlitteratur.

Syme, S. L. (1996) To prevent disease. The need for a new approach, in Blane, D., Brunner, E. & Wilkinson, R.G. (eds) *Health and Social Organization. Towards a Health policy for the 21st Century*. London and New York: Routledge: 21-31.

Syme, S. L. (1998) Social and economic disparities in health: thoughts about intervention, *Millbank Quarterly*, 76 (3): 493-506.

Syme, S.L. & Berkman, L.F. (1976) Social class, susceptibility and sickness, *American Journal of Epidemiology*, 104 (1): 1-8.

Tarlov, A.R. (1996) Social determinants of health: the sociobiological translation, in Blane, D., Brunner, E. & Wilkinson, R.G. (eds) *Health and Social Organization. Towards a Health policy for the 21st Century*. London and New York: Routledge: 71-93.

Thürmer, H. (1993) *Risk factors for, and 13-year mortality from cardiovascular disease by socioeconomic status*. ISM skriftserie nr. 26. Tromsø: Universitetet i Tromsø, Institutt for samfunnsmedisin.

Timms, D.W.G. (1996) Social mobility and mental health in a Swedish cohort, *Social Psychiatry and Psychiatric Epidemiology*, 31 (1): 38-48.

Timms, D. (1998) Gender, social mobility and psychiatric diagnoses, *Social Science & Medicine*, 46 (9): 1235-1247.

Townsend, P. (1990) Individual or social responsibility for premature death? Current controversies in the British debate about health, *International Journal of Health Services*, 20 (3): 373-392.

Townsend, P. & Davidson, N. (1982) *Inequalities in Health. The Black Report*, Harmondsworth: Penguin Books.

Turner, R.J. & Wagenfeld, M.O. (1967) Occupational mobility and schizophrenia: an assessment of the social causation and social selection hypotheses, *American Sociological Review*, 32 (February): 104-113.

Turner, R. J., Wheaton, B., and Lloyd, D.A. (1995) The epidemiology of social stress, *American Sociological Review*, 60 (February): 104-125.

Uitenbroek, D.G., Kerekovska, A. & Festchieva, N. (1996) Health life-style behavior and sociodemographic characteristics - A study of Varna, Glasgow and Edinburgh, *Social Science & Medicine*, 43 (3): 367-377.

Vatten, L. (1999) Brystkreft – risikofaktorer og mulighet for å forebygge, in NOU 1999:13 Kvinners helse i Norge, Oslo: Statens forvaltningstjeneste: 381-384.

Vågerö, D. (1992) Health inequalities from the cradle to the grave? Hypotheses on health policy, *International Journal of Health Sciences*, 3 (3/4): 175-183.

Vågerö, D. & Illsley, R. (1995) Explaining health inequalities: Beyond Black and Barker. A discussion of some of issues emerging in the decade following the Black Report. *European Sociological Review,* 11 (3): 219-241.

Wadsworth, M. E. J. (1986) Serious illness in childhood and its associations with later-life achievement, in Wilkinson, R. G. (ed) *Class and Health. Research and longitudinal data.* London: Tavistock Publications Ltd: 50-74.

Wadsworth, M. (1996) Family and education as determinants of health, in Blane, D., Brunner, E. & Wilkinson, R.G. (eds) *Health and Social Organization. Towards a Health policy for the 21st Century*. London and New York: Routledge: 152-170.

Wadsworth, M.E.J. (1997) Health inequalities in the life course perspective, *Social Science & Medicine*, 44 (6): 859-869.

Wadsworth, M.E.J. & Kuh, D.J.L. (1997) Childhood influences on adult health: a review of recent work from the British 1946 national birth cohort study, the MRC National Survey of Health and Development, *Paedriatic and Perinatal Epidemiology*, 11: 2-20.

Wagstaff, A., Paci, P. & Van Dorslaer, E. (1991) On the measurements of inequalities in health, *Social Science & Medicine*, 33(5): 545-557.

Waldmann, R. J. (1992) Income distribution and infant mortality, *Quarterly Journal of Economics*, 107: 1283-1302.

Warriner, C. K. (1981) Levels in the study of social structure, in Blau, P.M. & Merton, R.K. (eds) *Continuities in Structural Inquiry*, London and Beverly Hills: Sage Publications:179-190.

Waters, M. (1994) Succession in the stratification system: A contribution to the 'Death of Class' debate, *International Sociology*, 9 (3): 295-312.

Wennemo, I. (1993) Infant mortality, public policy and inequality – a comparison of 18 industrialized countries, *Sociology of Health & Illness*, 15 (4): 429-446.

West, P. (1988) Inequalities? Social class differentials in health in British youth, *Social Science & Medicine*, 27: 291-296.

West, P. (1991) Rethinking the health selection explanation for health inequalities, *Social Science & Medicine*, 32: 373-384.

West, P. (1997) Health inequalities in the early years: Is there equalisation in youth? *Social Science & Medicine*, 44: 373-384.

Whitehead, M. (1992) Concepts of equity and health, *International Journal of Health Services*, 22 (3): 429-445.

Whitehead, M. (1998) A European perspective on health inequalities, *Millbank Quarterly*, 76 (3): 469-492.

Wildawsky, A. (1977) Doing better and feeling worse: The political pathology of health policy, *Dædalus*, 106 (1).

Williams, S.J. (1995) Theorising class, health and lifestyles: can Bourdieu help us? *Sociology of Health & Illness*, 17 (5): 577-604.

Williams, S.J. (1998) 'Capitalising' on emotions? Rethinking the inequalities in health debate, *Sociology*, 32 (1): 121-139.

Wilkinson, R.G. (1986a) Socio-economic differences in mortality: interpreting the data on their size and trends, in Wilkinson, R. G. (ed) *Class and Health. Research and longitudinal data.* London: Tavistock Publications Ltd: 1-20.

Wilkinson, R.G. (1986b) Occupational class, selection and inequalities in health: a reply to Raymond Illsley, *Quarterly Journal of Social Affairs*, 2 (4): 415-422.

Wilkinson, R.G. (1987) A rejoinder to Illsley, *Quarterly Journal of Social Affairs*, 3 (3): 225-228.

Wilkinson, R.G. (1993) The impact of income inequality on life expectancy, in Platt, S., Thomas, H., Scott, S. & Williams, G. (eds) *Locating health. Sociological & Historical Explorations*. Aldershot: Avebury: 7-28.

Wilkinson, R. G. (1994) The epidemiological transition: from material scarcity to social disadvantage, *Dædalus*, 123 (4): 61-77.

Wilkinson, R. G. (1996) *Unhealthy Societies. The Afflictions of Inequality*. London and New York: Routledge.

Wilkinson, R.G. (1997) Health inequalities: relative or absolute material standards, *British Medical Journal*, 314 (22 February): 591-595.

Wilkinson, R. G. (1999) Income inequality, social cohesion, and health: clarifying the theory – a reply to Muntaner and Lynch, *International Journal of Health Services*, 29 (3): 525-543.

Wilkinson, R.G., Kawachi, I., & Kennedy, B.P. (1998) Mortality, the social environment, crime and violence, *Sociology of Health & Illness*, 20 (5):578-597.

Wingate, P. (1979) *The Penguin Medical Encyclopedia.* Harmondsworth: Penguin Books Ltd.

Wohlfarth, T. (1997) Socioeconomic inequality and psychopathology: are socioeconomic status and social class interchangeable? *Social Science & Medicine*, 45 (3): 399-410.

Wood, E.M. (1986) *The retreat from class. A new "true" socialism.* London: Verso.

Wright, E.O. (1985) *Classes*. London: Verso.

Waaler, H. T. (1984) Height, weight and mortality. The Norwegian experience, *Acta Medica Scandinavia* (suppl. 679): 1-56.

Zola, I.K. (1966) Culture and symptoms – an analysis of patients presenting complaints, *American Sociological Review*, 31: 615-630.

Aarø, L.E. (1986) *Health behaviour and socioeconomic status. A survey among the adult population in Norway*. Doctoral thesis. Bergen, Norway: University of Bergen, Department of Social Psychology.

Aase, A. & Storm-Furru, I. (1996) *Helse – helsetilstand – helsetjenester. Nasjonalatlas for Norge*. Hønefoss: Statens kartverk.

FIVE PAPERS

Paper I:

"How large are the differences – really? Self-reported long-standing illness among working class and middle class men", *Sociology of Health & Illness*, 18 (4): 475-498, 1996

Sociology of Health & Illness Vol. 18 No. 4 1996 ISSN 0141–9889, pp. 475–498

How large are the differences – really? Self-reported long-standing illness among working class and middle class men

Jon Ivar Elstad

Institute of Applied Social Research, Oslo

Abstract Social inequalities in health are regularly found in health interview surveys, but it has been doubted whether respondents' self-reported morbidity can accurately describe social class differences in health. Two hypotheses are current: members of the working class are commonly more preoccupied with their illnesses, and overreport them, compared with those of the middle class: secondly, the latter are more sensitive to their illnesses, have a better chance of getting them diagnosed, and consequently claim a wider array of illnesses in comparison with the former. This paper analyses answers to questions about long-standing illness from working class and middle class men, aged 25–64, in Norwegian surveys during the 1980s. Working class respondents experience their illnesses as more serious. An analysis of diagnostic categories of the illnesses of both sets of respondents indicates that this subjective experience is realistic. The hypothesis that a high level of psychological malaise among working class men produces a tendency to somatise psycho-social stress is not supported. The paper concludes that social inequalities in morbidity as shown by self-reported long-standing illness will often underestimate true class differences.

Key words: self-reported morbidity, health inequalities, long-standing illness, survey methodology, social class.

Introduction

Social inequalities in health are often found in health interview surveys (World Health Organization 1986). It has been doubted, however,

whether respondents' self-reported ill-health accurately measures social class differences in health (Aiach and Curtis 1990). People's opinion about their health is not only influenced by objective pathology and random individual idiosyncracies, but also by location in the social and cultural structure (Mechanic 1978:261, Fitzpatrick 1984). Culture has been found to influence ethnic variation in response to symptoms (Zola 1966, Bates and Rankin-Hill 1994), as well as gender differences in illness (Popay *et al.* 1993, Bendelow 1993). Cultural factors may also be involved when social class inequalities in health are described in surveys. Social classes may perceive, acknowledge, and communicate their health troubles differently. Measures of self-reported morbidity are influenced by illness behaviour and people's interaction with the health services (Blane *et al.* 1993). If such factors vary with class, social inequalities in health will be erroneously described by means of health interview surveys.

This paper will examine this topic further. The first sections review earlier research. Some studies claim that working class people tend to overreport their illnesses, while others contend the opposite. Explanations of such patterns and possibilities of historical change are considered. After that, some remarks on method introduce an empirical study of response patterns among working class and middle class men in four Norwegian interview surveys 1980–91. The study focuses on a common indicator of morbidity: long-standing illnesses. The analyses examine how self-reported long-standing illness acts as an indicator of social inequalities in health, and attempt to provide a better understanding of the magnitude of class differences in morbidity. By implication, the study also suggests some possible variations between social classes in illness and help-seeking behaviour in Norway during the 1980s.

Absolute and relative under/overreporting

Health problems may be overreported in interviews if respondents claim groundlessly that they have an illness; or underreported if conditions recognised as health problems both by respondents and doctors nevertheless remain unreported. If *absolute* overreporting or underreporting in this sense often occurs in surveys, their ability to describe health inequalities can be doubted. However, more relevant for the topic discussed in this paper is *relative* over/underreporting.

The basis for relative over/underreporting lies partly in the nature of health problems. There is often no unambiguous answer to the question whether a person is ill or not. One problem is divergent medical opinions. For example, to determine whether a person suffers from a chronic disease is often difficult because of 'a lack of clear onset criteria' (Blane *et al.* 1993:68). Another problem is the distinction between disease (medically defined disorders) and illness (lay defined health problems), which

may lead to disagreements between professional and lay judgements as to a person's health. Also when illness is in focus, as in this paper, it is often difficult to determine whether a person is ill or not. The so-called 'illness iceberg' implies that in any random population sample, a majority have some complaint, symptom or abnormality (Gerhardt 1989:260, *see also* Armstrong 1995). A large number of these illnesses are relatively non-serious. Thus, a person may have a condition which in principle can be given a medical diagnosis, but there is no compelling need for the person to declare him- or herself as ill and to report the disorder in an interview.

The existence of a large pool of non-serious illnesses is accompanied by the prominent role of social factors in determining whether ill-health is reported in health interviews. Blane *et al.* list subcultural expectations of health, structure of lay referral systems, interaction between symptoms and other aspects of a person's life, quality of locally available medical care, and the type of doctor-patient relationship as important in this respect (1993: 66). Conditions which have a sinister prognosis, are painful, and restrict one's activity, tend to be reported, but comparisons between reporting and medical records indicate that many diagnosed disorders are not reported in surveys (Mechanic 1978:147–8, *see also* Blaxter 1990:39). Symptoms and illnesses which have never been diagnosed are to a large extent non mentioned (Blaxter 1990:43–6), and even diagnosed illnesses are often forgotten or avoided in health interview surveys.

The indeterminate nature of illness, and the significant part played by illness behaviour in recognising illness, open up the possibility that social classes systematically apply different criteria when illnesses are reported. Lines separating illness from non-illness are not absolute, but dependent on prevailing lay and professional cultures, and predominant health cultures may differ between social classes. Therefore, one social class may regularly include a wider range of health troubles in comparison with another social class, both in terms of different types of problems, and in terms of more trivial and inconsequential disorders. To denote this situation, the terms 'relative overreporting' and 'relative underreporting' are used in this paper. They describe reporting behaviour of one social class in comparison with another, but imply no judgement as to whether the one or the other type of reporting behaviour is the more legitimate.

Hypotheses about relative overreporting among lower social strata

The hypothesis that lower social strata overreport their illnesses, compared with the middle classes, has been substantiated by claiming that the former are characterised by particular psychological patterns. Thus, Kadushin claimed in an early article (1964) that lower classes 'react more violently to [a medical condition] and are more concerned about it', 'are more fearful of serious diseases' and 'more likely to express their emotional anxieties in physical forms', and are 'less confident of their ability to cope successfully with serious illness'. He suggested that the explanation

for these tendencies was that among those who lack both material and cultural resources – money, knowledge, friends or relatives who are physicians, etc. – the outbreak of health difficulties may appear disproportionately dangerous. Therefore, their handling of such problems may often be inadequate, and their mental reactions will typically be exaggerated as compared with more resourceful social strata, who are more able to cope with their diseases.

The notion that poverty is associated with an intensified experience of illness has also been expressed in terms of the so-called 'culture of poverty' characterised by 'a sense of powerlessness, passivity, and fatalism' (Fitzpatrick 1984:58). The 'culture of poverty' syndrome is perhaps more suited to the extreme impoverishment in ghettos and backward rural areas, but some part of it may also be relevant among relatively secure sections of the working class in the West, for instance, fatalism: members of the working class may believe that they have less control over their lives and that neither preventative care nor their own life style can hinder diseases (Pill and Stott 1982, Fitzpatrick 1984:61–5, Davison *et al.* 1992). They tend to become resigned when falling ill, and are inclined to regard illness as an unavoidable part of their being, as compared to social classes who have a more activist attitude towards their illnesses.

This view is congruent with observations that diseases more easily develop into chronic illnesses among the working class. Haber observed that 'Some studies in Britain have found that the emotional problems of older working men are more important factors in hard-core sickness absence than the original physical condition' (1970:180). Attitudes of resignation, in combination, perhaps, with conscious or unconscious wishes to escape from alienated work, may lead to working class persons more rapidly integrating illness as part of their personal identity.

A related idea is that psychosomatic illnesses are widespread among the working class (Hollingshead and Redlich 1958; for references *see* Barsky and Klerman 1983:280, Larson and Marcer 1984, Blair 1986). In this context, psychosomatic illness can be defined as physical symptoms which suggest a somatic disease, although no relevant biomedical pathology can be found. According to this hypothesis, the high level of somatic illnesses among lower social classes, reflects a tendency to somatise their psychosocial problems. An elaborated hypothesis which leads to similar conclusions has been developed by Blair (1986). Based on Bernstein's theories about class differences in linguistic codes, Blair contends that the working class, due to life circumstances in terms of, for instance, 'child-rearing practices and educational and work activities', tends to be 'more likely to articulate distress through the more immediate site of the body' (1986:352). The 'elaborated code' prevalent among middle classes sustains a more personalised and differentiated notion of self, which corresponds to a preoccupation with the psyche. In contrast, members of the working class develop identities and attitudes to the world on the basis of

a 'restricted' vocabulary, leading to a tendency to express distress as bodily phenomena, for instance, in terms of somatic illness.

A somewhat different hypothesis is deduced from the theories of Parsons (1958) and Merton (1970[1938]), perhaps particularly tailored to the American scene. In their theories, success in terms of wealth, prestige and power is an overriding cultural value. Those who are not attaining it may be in need of legitimating their 'failure', and a possible option is to withdraw into the sick role, which exempts them from normal role obligations. The implication of this theory is that low status persons who experience their lives as failures, will – consciously or unconsciously (note that the sick role concept as envisaged by Parsons includes a motivational factor, *see* Gerhardt 1989:37–48) – tend to dramatise their health problems, in order to legitimate their position and retain their self-respect. In line with this, Cole and Lejeune (1972) studied self-defined health among American (often black) women who received various kinds of welfare support, and found evidence that those who expected no escape from their status as welfare recipients, tended to construct an overstated image of their ill-health.

Hypotheses about relative overreporting among higher social strata

An old formulation of the opposite hypothesis – that relative overreporting of health difficulties characterises the middle classes – focuses on the contrast between the coarse and rough life circumstances of the working class, and the refined and sophisticated ways of the middle class. Pain and other symptoms which prompt worries and help-seeking behaviour among the middle class will often be neglected among the working class. Such symptoms are considered part of everyday life. They accompany hard manual work and harsh conditions of living, and do not lead to any behavioural changes such as resting in bed or visiting a doctor. Two relevant studies are Koos's classic investigation of health and illness in an American town (1954), which found that 'upper-class persons were more likely than lower-class persons to view themselves as ill when they had particular symptoms' (referred to by Mechanic 1978:261), and Blaxter and Paterson's findings that working class women accepted as normal quite considerable levels of symptoms, for instance, recurrent ear infections among children (1982).

Another notion emphasises that knowledge of and relations with the health services are important when symptoms are recognised as illnesses and given a medical diagnosis. Better educated people could be expected to be better informed about the possible meaning of symptoms, and to have a more adequate vocabulary for describing their experiences. Moreover, there is probably a cultural affinity between health service professionals and middle class patients (Friedson 1970, *c.f.* Blane *et al.* 1993:67). Therefore, higher social strata may communicate more adequately with physicians (Fitzpatrick and Scambler 1984:60), and their

resources in terms of money and acquaintances enable them to demand high quality from their service providers. In consequence, it could be expected that a relatively larger part of the 'illness iceberg' among the middle class is acknowledged and given a diagnosis, and subsequently reported in health surveys.

Some attention has been given to the possibility of different concepts of health prevailing in middle class and working class cultures, which could influence the way health phenomena are perceived and communicated. A French analysis (d'Houtaud and Field 1984) indicates that members of the middle class conceive health in terms of personal vitality and enjoyments to a larger degree than those of the working class, who conceive it as being free from disabling diseases and enabling them to hold a job. One possible interpretation is that members of the middle class place higher demands on what constitutes health, and this could imply that smaller and more trivial disorders are considered problematic and therefore reported.

The existence of class-specific concepts of health is not beyond doubt, however, and Blaxter's analysis of the British Health and Lifestyle Survey (1990:13–35) does above all indicate that lay concepts of health are multidimensional in every social class, and tendencies to class profiles have unclear consequences for reporting in health interviews. Blane *et al.* suppose, however, that 'Subcultural expectations [of health] . . . may raise the perception of illhealth among the middle classes'. (1993:68) The reason for this may be different evaluations of the part played by one's own behaviour in creating and sustaining good health. The idea that health outcomes can be controlled by personal efforts, and that one should consciously mould one's behaviour in order to avoid health problems, is generally found to be more widespread among middle classes (Osler 1993, Aro *et al.* 1986, Aaro 1986, Blaxter 1990: Ch. 7). Greater value is placed on health, and this may be accompanied by an intensified awareness of minor symptoms. Furthermore, the finding that middle class people believe more than those of the working class that medical advances, knowledge and education have generally improved people's health (Blaxter 1990:155) could imply that they put greater emphasis on getting symptoms examined, and in consequence a wider array of disorders are diagnosed and later reported in health interviews.

Empirical evidence

The empirical evidence is scattered, and relevant studies use different data and methods, thus making comparisons difficult. Nevertheless, as a whole they seem to give only little support to the notion that working classes overreport, and somewhat more support to the opposite hypothesis. Studies have found a tendency for lower social strata to somatise life difficulties (*see*, for instance, the early and highly influential study of

Hollingshead and Redlich, 1958 as well as a newer and smaller study, Lenzi *et al.* 1993). In line with this, Blair (1986) found that workers were more inclined to explain illness (examples are cancer and depression) by physical factors.

Others, however, doubt that working class people generally overreport. American studies of self-reported illness compared with medical records have found no consistent differences in adequacy of reporting when respondents are classified according to educational or income level (Mechanic 1978:150). McBroom (1970) compared self-reported health and medical examinations among white Americans applying for welfare benefits, and concluded that 'there is no evidence to indicate that lower-status individuals tend to overreact to or overreport their symptomatology'. A Norwegian study found that undiagnosed symptoms suggesting cardiovascular disease were more frequent among working class than among middle class men (Thürmer 1993:94), and this could imply that underreporting is more frequent among the working class.

Similar conclusions are drawn from several British studies. Larson and Marcer (1984) tested the experience of pain among patients from different social classes, and concluded that '[the idea] that patients from the lower social classes have a higher perception of pain and are more neurotic than other groups is a myth, probably resulting from poor communication between clinicians and patients of dissimilar socioeconomic class'. Blaxter has concluded that 'the least advantaged tend to underestimate their ill-health' (1989:226), and her analysis of the Health and Lifestyle Survey led to the observation that 'the better-educated may give more informative answers' and that 'silent diseases' (disease without functional limitations) are reported more often by members of middle class than by those of the working class (1990:40,45). O'Donnel and Propper have examined the association between self-reported morbidity and physiological measures in the Health and Lifestyle Survey, and concluded that 'For a given level of self reported morbidity, individuals in the lower income groups are more likely to suffer multiple and more serious conditions and their health status, as measured by respiratory and cognitive functioning, is likely to be lower' (1991:17). The review by Blane *et al.* (1993) implies that most self-reported morbidity measures will particularly underestimate ill-health among the working classes. The exception is restricted activity, where greater physical demands in the occupations of manual workers are assumed to lead to a wider inclusion of illnesses as limiting.

The possibility of historical change

As we have seen, the majority of relevant empirical studies indicate that higher social strata, rather than lower social classes, overreport their illnesses. Factors which shape such patterns – class-specific cognitive

processes, vocabularies, values, illness behaviour, etc. – need not be stable across time or countries, however. Some interpretations of the present historical phase contend that class disparities have become less prominent in contemporary Western societies. According to Clark and Lipset (1991), distinct class cultures have more or less collapsed in the most advanced countries, and theories about late modernity (Giddens 1992) suggest that class-specific cultures have been unable to resist the ever-growing ideological pressure from national and global institutions.

A case in point may be Norway, with a relatively high standard of living, a relatively egalitarian income distribution (Ringen and Uusitalo 1992), and influential institutions which may have overcome class-specific health cultures. Assuming that the national educational system has provided similar health information to pupils from all social classes for some decades; that health professions, through their frequent contacts with patients, have educated individuals from all social strata more or less uniformly about health matters; and that the mass media have disseminated the same type of health information to a broad public for a long time, a possible hypothesis might be that only small differences between social classes persist in the way they report health problems.

Method

The literature reviewed above points towards three contrasting hypotheses: that lower social strata include more types of illnesses than higher social classes when they report their health problems; that – on the contrary – the latter overreport in comparison with the former; and that class-specific cultures have declined to the point where class differences in the way illnesses are reported in interviews have practically disappeared. The following empirical analysis will discuss the plausibility of these hypotheses, by studying how social classes in Norway during the 1980s reported long-standing illness in surveys.

Long-standing illness is in focus because of its centrality when social inequalities in morbidity are described. Other indicators, such as restricted activity, acute symptoms, or self-assessed health status, are also employed to monitor inequalities in health, but long-standing, or chronic, illnesses regularly have an important position – for instance in reports on social indicators in Norway (Central Bureau of Statistics 1992:124), Sweden (Vogel *et al.* 1988:272), Denmark (Bunnage 1992:80), Finland (Sauli *et al.* 1989: Table 5.1), the Netherlands (Netherlands Central Bureau of Statistics 1992:24,26) and Britain (Foster *et al.* 1995:86). Its centrality derives from the belief that information about long-standing illnesses is in a sense more objective than information about restricted activity, for instance, or self-assessed general health status (Aiach and Curtis 1990:268). The latter indicators are undoubtedly dependent on

subjective evaluations, while reporting of long-standing illness usually consists of medical diagnoses obtained from a physician (Blane *et al.* 1993:72). By stressing chronicity instead of illnesses in general, it is believed that the more serious conditions are registered (Blaxter 1989:207).

The method employed here is to examine the reporting of long-standing illnesses by means of other answers given in the same surveys. Thus, the approach consists of discussing how information about long-standing illnesses should be interpreted in the light of other answers given by respondents about diagnoses, duration of illnesses, impact on activity, and so on.

It can be objected that this method attempts to solve questions which basically concern respondents' subjectivity by referring to other data equally dependent on that same subjectivity. The view taken here, however, is that respondents' statements about long-standing illness and other health matters are generally valid in the sense that they match respondents' own opinions. Entirely unfounded fabrication of illnesses and deliberate suppression of relevant conditions, will of course occur in interviews, but it is assumed that such instances are rare and will not influence results in any consequential way. The issue is not whether social classes vary in their ability to provide valid answers. Rather, the issue is whether social classes are subject to different social forces and apply different criteria when they give their, by and large, equally valid answers. Accordingly, when a person reports a long-standing illness, the method employed here is not to doubt this information, but to utilise supplementary information given in the same interview to clarify the content of what has been reported.

When information obtained in health interviews is evaluated, the generally preferred method is to compare answers with supposedly more valid data: physicians' opinions, diagnostic tests, physiological measures, or other medically obtained information. Clearly, such comparisons are often very useful – the studies of McBroom (1970) and O'Donnel and Propper (1991) are relevant examples. Medically obtained information is unfortunately not available for the samples analysed here, and therefore the following analyses rely entirely on examinations of survey answers. It should be noted, however, that medically obtained information is not free from subjectivity, and the claim that medical judgement has a higher epistemological status than lay opinion can be questioned. Information from respondents' regular physicians is influenced by respondents' willingness to visit their doctor. Health examinations on random samples reduce this problem, but physicians' assessments are still often based upon what the patient says. Moreover, patients with chronic illnesses frequently acquire a high degree of expertise about their own conditions, and lay respondents' self-perceived health is often of better prognostic value than predictions made by physicians (Idler 1992). Such observations indicate the value of respondents' own opinions when studying health and illness.

Data

The Norwegian Central Bureau of Statistics conducted four Surveys of Level of Living in 1980, 1983, 1987 and 1991 (Central Bureau of Statistics 1982, 1985, 1988, 1992), interviewing a nationally representative sample aged 16–79. Long-standing illnesses were recorded by the question: 'Do you have any disease or disorder of a long-standing nature, any condition caused by accidents, or any handicap? Do report all such conditions, even if you consider them quite trivial' (author's translation). Three more questions ('Can you describe this in more detail? What did the doctor say it was? Which part of your body is afflicted?') were used to obtain further information. No definition was given of the term 'long-standing', and respondents had to decide themselves whether their illnesses qualified in this respect. After these questions had been answered, either by 'no' or by mentioning one or more illnesses, the question about long-standing illnesses was repeated, now accompanied by a check-list naming about forty common chronic diseases. The data files from the 1987 and 1991 surveys registered whether the respondent's illnesses were reported before or after the presentation of the check-list, but this was unfortunately not done in 1980 and 1983.

All reported illnesses were coded in terms of the three-digit diagnostic categories of the WHO's International Classification of Disease (ICD). The 8th revision (Central Bureau of Statistics 1973) was used in 1980–1987 and the almost identical 9th revision (Statistics Norway 1993) in 1991. When respondents had described their illnesses by means of diagnoses corresponding directly to the three-digit ICD categories, coding was usually straightforward. Otherwise, the codifiers (usually physicians) used all supplementary information registered by the interviewers in order to classify the illness.

Respondents who had mentioned at least one long-standing illness were asked: 'Do(es) your illness(es)/handicap(s) limit in any way your working capacity (paid employment, housework, education)?' Respondents were further asked three questions about symptoms during the last 6 months: palpitation, anxiety and nervousness, and depression. These questions are used later in this paper to divide the sample into 'afflicted with' (those who report at least one of these symptoms) and 'not afflicted with' psychological malaise.

The following analyses are restricted to male respondents, aged 25–64, currently employed either as workers (unskilled or skilled) or in white-collar occupations (lower, medium, and higher white collar). Occupations have been coded according to the Norwegian Central Bureau of Statistics' standard classification (Central Bureau of Statistics 1984), which resembles the Occupational Classes of the British Registrar General. Apart from the self-employed professions (lawyers, physicians, engineers etc.), who are classified as higher white collar, self-employed respondents are excluded from the following analyses – they are difficult

to place in the social hierarchy because their firms differ widely in size. Analyses were restricted to men, both in order to simplify and in order to avoid the troublesome question about the class location of women (Goldthorpe 1983, Dale *et al.* 1985, Dahl 1991). It is assumed that class-specific ways of reporting illness can be legitimately studied separately for each gender.

Table 1 describes the samples and their reporting of long-standing illness. The occupational distribution changed gradually during the 80s: the percentage of workers compared to white collar occupations diminished. Differences in long-standing illnesses between these occupational classes were quite stable, however. Therefore, it seems legitimate to pool the four surveys and analyse data as if it were one large survey, representing the average situation of the 1980s. Furthermore, respondents are divided into *working class* (unskilled and skilled workers) and *middle class* (lower, middle and higher white collar). This dichotomisation conceals differences within both the working class and the middle class. However, in order to investigate possible class-specific patterns of reporting, a broad division into a lower and a higher part of the social hierarchy seems tenable. Of course, a main reason for collapsing the samples and the categories is to obtain more precise estimates.

The following analyses are based on answers from 3720 male respondents, classified as working class (n = 1679) or middle class (n = 2041). Workers' average age was 42.9 years (SD = 11.6), as compared to 42.3 among middle class respondents (SD = 10.8). These small age differences can hardly have any notable effect on results. Nevertheless, in some of

Table 1: *Percentages reporting long-standing illness. Selected occupational categories*, men aged 25 to 64. Survys of Level of Living, Norway, 1980–83–87–91, separately and pooled.*

	Workers		*Middle class (white collar)*		
	Unskilled	*Skilled*	*Middle*	*Higher*	*Total*
1980	49.3	45.5	34.2	35.4	41.3
(N)	(306)	(167)	(342)	(127)	(942)
1983	50.8	44.5	39.6	38.9	43.6
(N)	(260)	(211)	(285)	(208)	(964)
1987	53.0	48.8	37.3	33.9	42.1
(N)	(200)	(166)	(367)	(186)	(919)
1991	51.4	46.7	41.2	34.7	43.2
(N)	(185)	(184)	(359)	(167)	(895)
Pooled	50.9	46.3	38.1	35.9	42.6
(N)	(951)	(728)	(1353)	(688)	(3720)

*Lower white collar has been combined with middle white collar because of small numbers.

the following analyses the samples of working and middle class respondents are standardised by age: a weighting procedure has made distributions identical with respect to 10-year-groups.

Some analyses use panel data from these surveys. In 1983, part of the 1980 sample was interviewed again, and this procedure was also followed in 1987 and 1991. Thus, the surveys have three two-wave panels: 1980–1983, 1983–1987, and 1987–1991, which have been pooled into one two-wave panel data set, with male respondents aged 25–64, and classified either as working class (n = 517) or middle class (n = 595) at the first interview. Among panel respondents the age distributions were practically identical – mean age 43.1 among working class (SD 11.6) and 43.0 (SD 11.3) among middle class respondents. By means of these panel data, changes over a three/four year period can be shown. Some respondents changed occupation between first and second interview, but in these analyses it seems permissible to classify respondents by occupation at first interview.[1]

Results

The experience of seriousness

Table 2 shows that long-standing illness was reported by 48.6 per cent of working class, as against 37.6 per cent of middle class respondents (age standardised results). The difference in percentages is 11.0 units, and in terms of odds ratio, the odds of long-standing illness afflicting a worker are 1.57 times the odds of a middle class respondent. Due to the large sample, this difference is of course highly significant in statistical terms, and it can also be judged to signify a substantial difference between these two social classes.

The following analyses will explore the credibility of this observed difference. Does it indicate the magnitude of class differentials realistically?

Table 2 shows that working class respondents who suffer from long-standing illness report, on average, slightly more conditions than middle class respondents in the same situation (1.49 as opposed to 1.39). This corresponds to respondents' opinion about how their illnesses limit their working capacity: among working class respondents who report long-standing illness, nearly half (46.6 per cent) claim that their illnesses interfere with their ability to work and/or to perform other activities, compared with 29.9 per cent among middle class respondents. Moreover, there is an interesting difference as to when conditions are reported in the interview. A larger proportion of illnesses reported by working class respondents was recorded immediately, while 'middle class illnesses' were more typically recorded after the check-list had been presented. Thus, it seems that working class respondents are more conscious of their illnesses than middle class respondents, and therefore need less stimulus in order to report them.

Table 2: *Reporting of long-standing illness, various aspects. Men, working class and middle class, aged 25 to 64, age standardised. Surveys of Level of Living, Norway, 1989–91, pooled.*

	Workers	Middle class	p-val
Long-standing illness (LSI) (%)	48.6	37.6	0.000
(Total no of respondents)	(1679)	(2041)	
Average No of LSI, among those who have LSI	1.49	1.39	0.014
Limiting LSI among those who have LSI (%)	46.6	29.9	0.000
(No of respondents with LSI)	(817)	(767)	
LSIs reported after check-list (%) * **	27.1	40.6	0.000
(No of LSIs reported 1987+1991)	(602)	(579)	
Panel data: Reporting LSI at second interview among those who reported LSI at first (%)*	74.8	64.6	0.020
(No of respondents reporting LSI at first interview)	(246)	(198)	

p-value based on chi square or analysis of variance
* not age standardised
** only 1987+1991

There are also class differences as to the permanence of long-standing illness. This is indicated by the panel data (lower part of Table 2). Among middle class respondents more than one-third of those suffering from long-standing illness at the first interview did not report long-standing illnesses some years later, as against one-fourth of working class respondents.[2]

Thus, the average working class respondent who says he suffers from long-standing illness, experiences this in distinctly different ways from the average middle class respondent. The working class respondent has more long-standing conditions, feels that these conditions are more likely to limit his working capacity, is more conscious of his illnesses, and claims more often that they persist over time.

An obvious interpretation of this pattern is of course that long-standing illnesses reported by working class respondents are, on average, *more serious*, than those reported by middle class respondents. Therefore, they have a larger impact (are felt more limiting), they come more ready to one's mind (they cannot so easily be forgotten, because of pain, prognosis, etc.), and they tend to persist (say, because their severity makes treatment ineffective). According to this interpretation, the middle class respondent commonly includes less serious conditions. If this is the case, then the difference observed in percentages afflicted with long-standing illness – 48.6 per cent as against 37.6 per cent – *underestimates* the true magnitude of social inequality.

There is, however, an *alternative* interpretation: when afflicted with dis-

ease, working class, compared with middle class men, tend *to define* their situation as more severe and critical. They adapt to this definition, both in terms of notions of self and in terms of behaviour. Thus, illness is more firmly incorporated into one's identity, experienced more consequentially in terms of limited working capacity, and considered less reversible.

Given that this alternative interpretation is correct, long-standing illnesses reported by the two social classes could be equally serious on average, although workers experience them as more severe. A more extreme variant of this alternative interpretation is that the 'original' difference (48.6 as opposed to 37.6 per cent) represents an inflated measure of inequalities in health, because working class respondents tend to include not only similar, but also less serious conditions, than do middle class.

Analysing seriousness by means of diagnostic categories

The plausibility of the 'obvious' as against the 'alternative' interpretation just described can be discussed further through an examination of the diagnostic categories recorded in these data, which can be used in order to estimate the seriousness of long-standing illnesses reported by the two social classes.

All reported illnesses were coded by means of three-digit ICD diagnostic categories. The alternative interpretation outlined above could imply that working class men tend to include less serious conditions *within* each diagnostic category. For example, when they report hypertension, the mean blood pressure would be lower, and when they report rheumatism, less inflammation would commonly be found. As objective data (for instance, physiological measurements of the type used by O'Donnel and Propper) is lacking, the average seriousness of illnesses within each diagnostic category cannot be examined here.

Data can be used, however, to compare the illness panorama of working class with that of middle class respondents. The hypothesis that working class respondents tend to report less serious conditions would be supported if their illness panorama included *a higher proportion of less serious illnesses*. The recorded diagnostic categories can shed light on this question.

In the data files, 224 different three-digit ICD categories have been used to code the 2299 cases of illnesses reported by the respondents. Many of these diagnostic categories are used only occasionally – for instance, 157 categories are employed in only one to five cases.

A handful of diagnostic categories occur frequently, however. For instance, among the total number of 2299 illnesses, 265 are coded as 'vertebrogenic pain syndromes', 168 as 'essential benign hypertension', and 165 as 'other eczema, dermatitis'. Twenty-one diagnostic categories are recorded 25 times of more in these data, and 1568 of the total of 2299 illnesses (68.2 per cent) are classified by means of these twenty-one diagnostic categories.

These categories can be termed 'common illnesses'. They occur frequently, and the corresponding disorders are widespread in the (male) population. These common illnesses have been classified according to presumed seriousness. Two experienced physicians have, independently of each other, evaluated the twenty-one diagnostic categories along three dimensions: (probability of) mortality, functional limitations, and need of medication/treatment. They were also told that patients were males, aged 25–64, and not permanently excluded from paid employment because of illness. There were of course some discrepancies, but also a basic agreement, which led to a classification of four diagnostic categories as probably serious illness, while thirteen diagnostic categories were classified as of moderate/varying seriousness, and four as probably less serious.[3]

Diagnostic categories which occur less frequently have not been classified in this way. Illnesses represented by a particular diagnostic category will often occur along the entire scale from light to grave conditions, and when a type of illness is recorded only occasionally, it is hard to guess its seriousness. When a type of illness occurs frequently, however, it can be expected that its distribution of seriousness in these data resembles the true distribution. Therefore, it is legitimate to classify common illnesses according to presumed seriousness, but this is less defensible for uncommon illnesses.

Table 3 describes the composition of long-standing illnesses reported by the two social classes. Common illnesses which are (probably) serious constitute about 7 per cent of illnesses reported by working class as well

Table 3: *Classification of long-standing illnesses reported by men aged 25–64, workers and middle class, Surveys of Level of Living, Norway, 1980–91, pooled.*

	Number of illnesses				*Illness per 1000 respondents**		
	Workers		*Middle Class*				
Type of illness	*No*	*%*	*NO*	*%*	*Worker*	*MidCl*	*Ratio*
Common illnesses							
– serious	85	6.9	76	7.1	51	37	1.36
– moderately serious	586	47.6	457	42.8	349	224	1.56
– less serious	169	13.7	196	18.4	101	96	1.05
All common illnesses	840	68.2	729	68.3	500	357	1.40
Medium common illnesses	158	12.8	143	13.4	94	70	1.33
Uncommon illnesses	110	8.9	113	10.6	66	55	1.19
Injuries ICD 800–999	124	10.1	82	7.7	74	40	1.84
All illnesses	1232	100.0	1067	100.0	734	523	1.40

Common illnesses = diagnostic categories recorded 25+ times. Medium common = recorded 6–24 times, uncommon illnesses = recorded 1–5 times. *Estimates of illnesses per 1000 respondents are rounded. Ratios workers'/middle class' illnesses per 1000 are calculated from unrounded figures.

as by middle class respondents. Common illnesses of moderate/varying seriousness are more frequent among workers (47.6 as opposed to 42.8 per cent), while common illnesses which are less serious are more frequent among middle class respondents.

Thus, the results do *not* indicate that working class tend to include less serious conditions, compared with middle class respondents. If anything, they point in the opposite direction: less serious illnesses constitute 18.4 per cent of 'middle class illnesses', as against 13.7 per cent of workers' illnesses.[4]

Other results displayed in Table 3 could also indicate a more inclusive manner of reporting among middle class respondents. The working-middle class ratio of illnesses per 1000 respondents increases from 1.19 for uncommon illnesses (diagnostic categories recorded 1–5 times), to 1.33 for medium common illnesses (6–24 times), and finally to 1.40 for common illnesses. Why are class differences apparently minimal with respect to rare types of illnesses? Their scarcity could very well imply that many of them remain undiagnosed. If middle class respondents receive better health services, there is less chance that such illnesses stay unacknowledged among this group. It can be expected that common illnesses are seldom overlooked by the doctor, because they are well known and therefore easier to diagnose. This may explain the perhaps puzzling pattern that class inequalities are particularly small with respect to uncommon illnesses. The implication of this is that underreporting of uncommon illnesses is particularly widespread among members of the working class.

Long-standing illness and psychological malaise

The topic of this paper can also be addressed from another angle by means of these data. As discussed in the introduction, some researchers have assumed that the working class overreport their illnesses in comparison with the middle class because of a particular association between psychological factors and the recognition of illness. It has been argued that working class people have a generally high level of anxiety, nervousness and similar psychological problems, which result in a disproportionately pessimistic account of the seriousness of their diseases. Moreover, they are assumed to somatise to a greater degree than members of the middle class: their psycho-social difficulties are to a higher degree expressed as somatic complaints.

Accordingly, a closer look at the association between psychological factors and reporting of long-standing illness is appropriate. This can be done by using the answers about psychological symptoms recorded in these data for dividing the sample into afflicted or not afflicted with psychological malaise.

The ideas described above can be specified in two ways. The first is based on two assumptions: that psychological malaise causes part of the experience of long-standing (somatic) illness, and that psychological

malaise occurs much more frequently among working class people. The combination of these two explains (part of) the higher level of long-standing illness among workers. This reasoning does not claim that the consequence of psychological malaise differs between social classes. Both working and middle class respondents suffering from it may transfer some of that distress to somatic symptoms. But as psychological malaise afflicts working class men more often, their long-standing illnesses can also more often be explained as a consequence of malaise.

Psychological malaise, as it is measured here, afflicts 24.0 per cent of working class, as against 17.5 per cent of middle class respondents (these percentages can be estimated from Table 4). The difference in reported malaise between these two social classes is therefore not particularly large. Percentages suffering from malaise in both social classes are much lower than percentages reporting long-standing illness. Thus, it does not seem credible that their high level of anxiety, nervousness and similar psychological problems account for the fact that working class respondents report more long-standing illness. Of course, it is possible that such problems are not very well measured in these data, and that they are especially avoided in the answers given by working class respondents. However, such underreporting would have to be very large if psychological malaise were to explain much of the class difference in long-standing illness. Therefore, the hypothesis that a particularly high level of psychological malaise among working class men explains the differences in reported long-standing illness is hardly borne out.

The second proposition does not necessarily assume that workers are more afflicted with psychological malaise, but that they react differently

Table 4: *Psychological malaise and long-standing illness (LSI). Men aged 25 to 64, workers and middle class. Surveys of Level of Living, Norway, 1980–91, pooled cross-sectional data and panel data.*

Psychological malaise	*Afflicted*			*Not afflicted*		
	Work	*MidCl*	*Total*	*Work*	*MidCl*	*Total*
Long-standing illness (%)	68.6	58.9	64.0	42.3	33.1	37.1
Limiting LSI (%)	45.3	26.8	36.6	15.6	7.9	11.2
(No of respondents)	(403)	(357)	(761)	(1276)	(1684)	(2959)
Proportion LSIs reported after checklist (%)	25.2	42.0	32.0	28.4	40.0	34.6
(No of LSIs 1987+1991)	(246)	(169)	(415)	(356)	(410)	(766)
Panel data						
Reporting LSI at second interview among those reporting LSI at first (%)	77.2	61.8	71.4	73.4	65.7	69.7
(No of respondents with LSI at first interview)	(92)	(55)	(147)	(154)	(143)	(297)

to it. Middle classes are thought to acknowledge psycho-social problems for what they are, while workers are more reluctant to explain their distress in terms of psychological problems, and tend to somatise them. Thus, psychological malaise will have a more pronounced tendency to coexist with somatic long-standing illness among workers than among middle class respondents. In statistical terms, this will appear as an interaction effect: workers afflicted with malaise will have a particularly high level of long-standing illness.[5]

Table 4 indicates, however, that psychological malaise has no particular effect on the reporting of long-standing illness among working class, different from the effect among middle class, respondents. Among those afflicted with malaise, more of the former (68.6 per cent) than the latter (58.9 per cent) report a long-standing illness.[6] This difference is however equal to the difference among those not afflicted with malaise (42.3 as against 33.1 per cent). As expected, malaise and long-standing illness are associated, but this association is exactly the same for working class and middle class respondents, and Table 4 shows no tendency that malaise among the former has a particularly strong effect on the reporting of long-standing illness.

Thus, these results do not support the hypothesis that the psychological problems of working class men can account to any large extent for their high level of self-reported long-standing illness. In passing, it can also be noted that Table 4 actually indicates that psychological malaise has hardly any effect on the experience of long-standing illness. If respondents afflicted with psychological malaise were particularly conscious of and preoccupied with their illnesses one would expect that they would have tended to report their long-standing illnesses at an early stage of the interview, *i.e.* before presentation of the check-list. However, Table 4 shows that the distribution of long-standing illnesses reported early and late in the interview is not influenced by respondents' malaise. Neither is the persistence of long-standing illness influenced in any clear way by reported psychological problems, *see* the lower part of the Table.

Conclusion

This paper has asked whether the true magnitude of inequalities in health is erroneously described by health interviews because of social class differences in the seriousness of illnesses which are reported. The hypothesis that class differences in ill-health appear unduly large when measured by self-reported morbidity, because of a tendency among the working class to include more trivial illnesses, was not supported by the data. Rather, the opposite idea appeared more credible. The experience of working class respondents who report that they suffer long-standing illness, compared with middle class in the same situation, indicates that their long-

standing illnesses are more serious: they are more conscious of their illnesses, consider them more limiting, and longer lasting than do middle class respondents.

This could mean that the description of inequalities in health by means of self-reported long-standing illness understates the magnitude of class differences. But the possibility that the serious nature of the experience is influenced by a tendency among working class men to exaggerate their health problems must be considered. However, an analysis of the illness panorama – *i.e.* the types of illnesses reported by working and middle class respondents respectively – indicates that workers commonly report a narrower range of illnesses. An analysis of diagnostic categories used to code reported illnesses shows that a larger proportion of illnesses reported from middle class respondents is milder and less serious. Furthermore, an analysis of the interplay between psychological malaise and reporting of long-standing illness shows no signs that the higher reporting of long-standing illnesses among workers is due to their anxiety and nervousness. Mental health problems may lead to somatisation, but these data do not indicate that this is particularly widespread among workers. Perhaps the most common hypothesis among researchers who believe that surveys of self-reported morbidity tend to overstate class inequalities is that the high level of long-standing (somatic) illness among the working class is explained by their high level of anxiety and their tendency to somatise psycho-social problems. This found no support in these data, and thus makes the hypothesis about working class overreporting less plausible.

Two main conclusions emerge from this paper. The first is that self-reported morbidity, at least in terms of percentages suffering from long-standing illness, probably underestimates the true class differentials in morbidity. The findings are of course bound to a particular social context – Norway in the 1980s – but it should be noted that the general pattern found here seems to correspond to several British analyses. Therefore, this may be a typical and widespread effect. Self-reported long-standing illness is regularly used in many social reports to indicate social inequalities in health. It seems plausible that these social reports often underestimate the extent of social inequalities, and this should be considered when such results are interpreted.

The second conclusion refers to the need for developing morbidity indicators which give a less biased picture of social inequalities, for use in health interviews. For obvious reasons this is problematic: reporting of health difficulties in interviews is strongly affected by cognition processes, predominant values, illness behaviour, and help-seeking behaviour. Norway could be expected to be a country with relatively small class differences; nevertheless, this paper suggests that in the 1980s, a set of social processes was operating which produced a higher sensitivity to health problems among middle classes. The exact mechanisms involved could

not be analysed here, but three factors could well be further investigated: healthism, medical knowledge, and relations with the health services. To overcome completely the distorting influence of class-specific cultures on self-reported morbidity is perhaps impossible, but improved indicators could possibly reduce the problem. Blane *et al.* (1993) have proposed that investigations should concentrate on specific age groups individually, and that efforts should be spent on developing sets of questions to elicit information about specific morbidity. I endorse this recommendation. It should, however, be underlined that validity testing is always necessary – note the perhaps surprising result (which corresponds to observations made by Aiach and Curtis 1990) that the probable effect of 'improving' survey instruments by including more questions and a check-list, was that middle class respondents to a higher degree than working class reported their more trivial and inconsequential long-standing illnesses. This resulted in an attenuated and less correct picture of the true magnitude of class differences in morbidity.

Address for correspondence: Jon Ivar Elstad, Institute of Applied Social Research, Munthesgt. 31, 0260 Oslo, Norway.

Acknowledgements

Data used in this paper have been collected and processed by Statistics Norway (formerly Central Bureau of Statistics, Norway), and made available for researchers by the Norwegian Social Science Data Service. Neither of these institutions is responsible for the analyses and interpretations made here. Discussions with Espen Dahl, Institute for Applied Social Science, Oslo, have been very helpful when writing this paper. Thanks to Susanne Rudberg, MD, PhD, St. Goran's Hospital, Stockholm, and Goran Wiklander, MD, Nacka Hospital Stockholm, for expert help in classifying diagnostic categories.

Notes

1 It should be noted that the representativeness of the pooled two-way panel sample is more questionable than usual in cross-sectional surveys. The second-time point observations in the 1980–83 panel are to some degree identical to the first-time point observations of the 1983–87 pane, and so on. Thus, the panel data are not freely sampled, and the variability is restricted. Furthermore, the panel samples are plagued, as most panel studies are, by sample attrition. Usually, the response rate is 76–79 per cent in these surveys (Central Bureau of Statistics 1982:14, 1985:14, 1988:14), but in the two-wave panel data, non-respondents accumulate, bringing the response rate down to about 66 per cent (1985:16, 1988:16).

2 That some long-standing illnesses apparently 'disappear' after some time has been observed in several panel studies – see for instance Wärneryd 1991 and

Swain 1993:51, 61. In part, this reflects ordinary problems of reliability, but there are also substantial reasons for this pattern. There are borderline illnesses which the respondent decides to report in one interview but neglects in the next. Moreover, it is not unusual that one recovers from some chronic diseases, and that some illnesses have passive phases with practically no impact. – The impression given in table 2 of workers' persistent long-standing illnesses in comparison with middle class respondents could be misleading if there was a tendency for workers to report *different* types of illnesses at each interview. However, when type of illness is controlled for, the same pattern of stability in workers' reporting, compared with middle class respondents, emerges.

3 These categories were classified as *serious*: ICD no 250 diabetes mellitus, ICD no 410 + 412 acute myocardial infarction + chronic ischaemic heart disease, 413 angina pectoris, and 493 asthma. Thirteen categories were classified as moderate/varying seriousness: 346 migraine, 389 deafness, hearing problems, 401 essential benign hypertension, 503 chronic sinusitis, 531–533 ulcer, stomach, duodenum + peptic ulcer, site unspecified, 712 rheumatoid arthritis etc., 713 osteo-arthritis, allied conditions, 717 other non-articular rheumatism, 718 rheumatism, unspecified, 725 displacement of intervertebral disc, 728 vertebrogenic pain syndromes, 731 synovitis, bursitis, etc., and 790 nervousness and debility. The following four categories were classified as *less serious*: 454 varicose veins, 507 hay fever, 692 other eczema, dermatitis, and 696 psoriasis and allied conditions.

4 It should be noted, however, that this result emerges primarily because of middle class respondents' frequent reporting of two commonly non-serious illnesses – varicose veins and hay fever.

5 The validity of this analysis presupposes that those who express their psychosocial troubles in terms of somatic complaints, will also admit in the interview that they suffer from psychological malaise. If mental health problems are completely suppressed and entirely transformed into somatic illness, this analysis would not be possible.

6 To be exact, this analysis should have focussed on *somatic* long-standing illness, in order to accentuate the association between malaise and somatisation. But as mental health problems (ICD codes 290–315 + 790) make up only 61 of 2299 reported long-standing illnesses (2.7 per cent), an elimination of mental health problems from the variable long-standing illness makes no effect on the results.

References

Aaro, L.E. (1986) *Health Behaviour and Socioeconomic Status. A Survey among the Adult Population in Norway*. PhD thesis, Bergen: University of Bergen.

Aiach, P. and Curtis, S. (1990) Social inequalities in self-reported morbidity: interpretation and comparison of data from Britain and France, *Social Science and Medicine*, 31, 267–74.

Armstrong, D. (1995) The rise of surveillance medicine, *Sociology of Health and Illness*, 17, 393–404.

Aro, S., Rasanen, L. and Telama, R. (1986) Social-class and changes in health-related habits in Finland in 1973–1983, *Scandinavian Journal of Social Medicine*, 14, 39–47.

Barsky, A.J. and Klerman, G.L. (1983) Overview – Hypochondriasis, bodily complaints, and somatic styles, *American Journal of Psychiatry*, 140, 273–83.

Bates, M.S. and Rankin-Hill, L. (1994) Control, culture and chronic pain, *Social Science and Medicine*, 39, 629–45.

Bendelow, G. (1993) Pain perceptions, emotions and gender, *Sociology of Health and Illness*, 15, 273–94.

Blair, A. (1986) *Locating Distress: Social class and the Contextualisation of Illness Experience*. PhD thesis, Brighton: University of Sussex.

Blane, D., Power, C. and Bartley, M. (1993) The measurement of morbidity in relation to social class. In Abel, T., Geyer, S., Gerhardt, U., Siegrist, J. and van den Heuvel, W. (eds) *Medical Sociology: Research on Chronic Illness*. Bonn: Informationszentrum Sozialwissenschaften.

Blaxter, M. (1989) A comparison of measures of inequality in morbidity. In Fox, J. (ed) *Health Inequalities in European Countries*. Aldershot: Gower.

Blaxter, M. (1990) *Health and Lifestyles*. London and New York: Routledge.

Blaxter, M. and Paterson, E. (1982) *Mothers and Daughters: a Three-generational Study of Health Attitudes and Behaviour*. London: Heinemann Educational Books.

Bunnage, D. (ed) (1992) *Living Conditions in Denmark. Compendium of Statistics 1992*. Copenhagen: Danmarks Statistik – Socialforskningsinstitutet.

Central Bureau of Statistics (1973) *Standard for Gruppering av Sykdommer – Skader – Dødsårsaker*. [Classification of Diseases – Injuries – Causes of Mortality.] Oslo: SSH 24.

Central Bureau of Statistics (1982) *Survey of Level of Living* 1980. Oslo – Kongsvinger. NOS B 320.

Central Bureau of Statistics (1984) *Standard for Inndeling Etter Sosioøkonomisk Status*. [Classification of Socioeconomic Status.] Oslo – Kongsvinger.

Central Bureau of Statistics (1985) *Survey of Level of Living 1983*. Oslo – Kongsvinger. NOS B 511.

Central Bureau of Statistics (1988) *Survey of Level of Living 1987*. Oslo – Kongsvinger. NOS B 772.

Central Bureau of Statistics (1992) *Survey of Level of Living 1991*. Oslo – Kongsvinger. NOS C 43.

Clark, T.N. and Lipset, S.M. (1991) Are social classes dying? *International Sociology*, 6, 397–410.

Cole, S. and Lejeune, R. (1972) Illness and the legitimation of failure, *American Sociological Review*, 37, 347–56.

Dahl, E. (1991) Inequality in health and the class position of women – the Norwegian experience, *Sociology of Health and Illness*, 13, 492–505.

Dale, A., Gilbert, N.G., and Arber, S. (1985) Integrating women into class theory, *Sociology*, 19, 384–408.

Davison, C., Frankel, S. and Smith, G.D. (1992) The limits of lifestyle: re-assessing 'fatalism' in the popular culture of illness prevention, *Social Science and Medicine*, 34, 675–85.

D'houtaud, A. and Field, M.G. (1984) the image of health – variations in perception by social-class in a French population, *Sociology of Health and Illness*, 6, 30–60.

Fitzpatrick, R. (1984) Lay concepts of illness. In Fitzpatrick, R. *et al.* (eds) *The Experience of Illness*. London: Tavistock Publications.

Fitzpatrick, R. and Scambler, G. (1984) Social class, ethnicity, and illness. In Fitzpatrick, R. *et al.* (eds) *The Experience of Illness*. London: Tavistock Publications.

Foster, K., Jackson, B., Thomas, M., Hunter, P. and Bennett, N. (1995) *1993 General Household Survey*. London: Office of Population Censuses and Surveys, Series GHS No. 24.

Freidson, E. (1970) *Profession of Medicine. A Study of Applied Knowledge*. New York: Harper & Row.

Gerhardt, U. (1989) *Ideas about Illness. An Intellectual and Political History of Medical Sociology*. London: Macmillan.

Giddens, A. (1992) *The Consequences of Modernity*. Cambridge: Polity Press.

Goldthorpe, J.H. (1983) Women and class analysis: in defence of the conventional view, *Sociology*, 17, 465–88.

Haber, L.D. (1970) Age and capacity devaluation, *Journal of Health and Social Behaviour*, 11, 167–82.

Hollingshead, A.B. and Redlich, F.C. (1958) *Social Class and Mental Illness*. New York: Wiley.

Idler, E.L. (1992) Self-assessed health and mortality: a review of studies, *International Review of Health Psychology*, 1, 33–54.

Kadushin, C. (1964) Social class and the experience of ill health, *Sociological Inquiry*, 34, 67–80.

Koos, E. (1954) *The Health of Regionville: What the People Thought and Did about It*. New York: Columbia University press.

Larson, A.G. and Marcer, D. (1984) The who and why of pain – analysis by social-class, *British Medical Journal*, 288, 883–6.

Lenzi, A., Lazzerini, F., Marazziti, D., Raffaelli, S., Rossi, G. and Cassano, G.B. (1993) Social class and mood disorders – clinical features. *Social Psychiatry and Psychiatric Epidemiology*, 28, 56–9.

McBroom, W. (1970) Illness, illness behavior and socioeconomic status, *Journal of Health and Social Behavior*, 11, 319–26.

Mechanic, D. (1978) *Medical Sociology*. Second Edition, New York: The Free Press.

Merton, R.K. (1970[1938]). Anomie and social structure. In Worsley, P. (ed) *Modern Sociology. Introductory Readings*. Harmondsworth: Penguin Books Ltd.

Netherlands Central Bureau of Statistics. (1992) *Netherlands Health Interview Surveys 1981–1991*. The Hague: SDU Publishers.

O'Donnel, O. and Propper, C. (1991) Equity and the distribution of UK National Health Service resources, *Journal of Health Economics*, 10, 1–19.

Osler, M. (1993) Social class and health behavior in Danish adults – a longitudinal study, *Public Health*, 107, 251–60.

Parsons, T. (1958) Definitions of heath and illness in the light of American values and social structure. In Jaco, E.G. (ed) *Patients, Physicians, and Illness*. Glencoe, Illinois: The Free Press.

Pill, R. and Stott, N.C.H. (1982) Concepts of illness causation and responsibility, *Social Science and Medicine*, 16, 42–51.

Popay, J., Bartley, M. and Owen, C. (1993). Gender inequalities in health – social position, affective-disorders and minor physical morbidity, *Social Science and Medicine*, 36, 21–32.

Ringen, S. and Uusitalo, H. (1992) Income distribution and redistribution in the

Nordic welfare states. In Kolberg, J.E. (ed) *The Study of Welfare States Regimes*. Armonk, New York: M.E. Sharpe.

Sauli, H., Ahola, A., Lahelma, E. and Savolainen J. (1989) *Living Conditions in Figures. The 1986 Survey on Living Conditions in Finland*. Helsinki: Central Statistical Office of Finland.

Statistics Norway (1993) *Klassifikasjon av Sykdommer, Skader og Dødårsaker. Norsk Utgave av ICD-9*. [Classification of Diseases, Injuries and Causes of Death. ICD-9, Norwegian version.] Oslo-Kongsvinger: SSB NOS Standarder.

Swain, V.J. (1993) Changes in self-reported health. In Cox, B.D., Huppert, F.A. and Whichelow, M.J. (eds) *The Health and Lifestyle Survey: Seven Years On*. Aldershot: Dartmouth.

Thürmer, H. (1993) *Risk Factors for, and 14-year Mortality from Cardiovascular Disease by Socioeconomic Status*. Tromsø: University of Tromsø, Institute of Community Medicine, ISM Skriftserie Nr. 26.

Vogel, J., Andersson, L.G., Davidsson, U. and Häll, L. (1988) *Inequality in Sweden. Trends and Current Situation*. Stockholm: Report No. 58, Statistics Sweden.

Wärneryd, B. (1991) *Levnadsförhållanden, Appendix 12. Återintervjustudie i Undersökningen av Levnadsförhållanden (ULF) 1989*. [Conditions of Living, Appendix 12. Study of Reinterviews in the Investigation of Conditions of Living 1989.] Stockholm: Statistiska Centralbyrån.

World Health Organization (1986) *The Health Burden of Social Inequities*. Copenhagen: WHO Regional Office for Europe.

Zola, I. (1966) Culture and symptoms: an analysis of patients presenting complaints, *American Scoiological Review*, 31, 615–30.

Paper II:

"The psychosocial perspective on social inequalities in health",
Sociology of Health & Illness, 20 (5): 598-618, 1998

Sociology of Health & Illness Vol. 20 No.5 1998 ISSN 0141–9889, pp. 598–618

The psycho-social perspective on social inequalities in health

Jon Ivar Elstad

NOVA – Norwegian Social Research, Oslo

Abstract Doubts about the viability of material explanations of social inequalities in health have led to a renewed focus on the aetiological role of psychological stress, and, moreover, on how psychological stress is generated by society's inequality structures. Some researchers maintain that the emerging psycho-social perspective will become the dominant paradigm in research on health inequalities. After commenting on some aetiological topics, the paper outlines how a comprehensive understanding of health inequalities can be constructed from the social stress model, the self-efficacy approach, the sociology of emotions, and the social cohesion approach. The emerging perspective is a striking attempt to deal with health inequalities, as it seems to solve some of the difficulties that other perspectives have had in accounting for existing empirical patterns. Nevertheless, it is perhaps too much to claim that it signifies a paradigm shift. It should rather be considered as an enrichment of the social causation explanation. The latter, when studying health
inequalities, should be developed further by considering both material and psycho-social environments, and their mutual interaction.

Keywords: health inequalities, psycho-social explanations, stress, emotions, self-efficacy, social cohesion

Introduction

Recent debates indicate that most researchers believe that genetics, social selection, and access to health services explain only minor parts of current social inequalities in health (Blane *et al.* 1996, Carroll *et al.* 1996, Evans 1994, Evans and Stoddart 1994, Tarlov 1996, Vågerö and Illsley 1995, Wilkinson 1996b: 53–71, Baird 1994, Blane *et al.* 1993). This renders the

social causation explanation more plausible. However, even the two main approaches within social causation, focusing on health-related behaviour and material standard of living, seem insufficient.

Health-damaging behaviour, such as smoking, unhealthy diets, and accident-prone behaviour, is usually overrepresented among lower social strata. This accounts, however, for only some part of current health variations (Marmot 1986, Blane *et al.* 1996: 5–7, Carroll *et al.* 1996: 29, Evans 1994, Wilkinson 1996b: 63–6). Moreover, as health-related behaviour is seldom chosen 'freely', but is heavily influenced by social status and cultural milieus, health-related behaviour can hardly be regarded as the basic cause of health inequalities.

Social medicine has primarily explained health variations by harmful physical factors in the environment, such as poor nutrition or polluted and hazardous workplaces. Scepticism has, however, grown as to the explanatory power of material explanations in contemporary Western societies. If material shortages were the major determinant of ill-health, one would expect that social inequalities in health would diminish as post-war improvements in standard of living raised the majority above certain threshold levels (Kadushin 1964, Blane *et al.* 1996: 3). But although life expectancy has increased, social inequalities in health have persisted (Carroll *et al.* 1996, Dahl and Kjærsgaard 1993, Link and Phelan 1995). If population health is determined by economic level, one would predict that each country's life expectancy would correspond to its international economic ranking. But Wilkinson's research (1990, 1994, 1996b: 72–109, 1997) indicates that life expectancy in affluent countries is associated with *income equality* rather than with average level of income, and his findings are echoed by others (Wennemo 1993, Kaplan *et al.* 1996, Kennedy *et al.* 1996, Waldmann 1992, Kawachi and Kennedy 1997). The so-called 'challenge of the gradient' (Adler *et al.* 1994) is another aspect of the puzzle, illustrated by findings of the Whitehall studies that not only low-level employees, but even civil servants close to the top, had worse health than the top category (Marmot 1986, Marmot 1994). Differences in current material standards could hardly account for these mortality differentials, as these two occupational groupings are both favourably placed in the income hierarchy.

The psycho-social perspective: towards a paradigm shift?

Thus, established explanations are challenged, and alternative interpretations have gained support, in a manner resembling Kuhn's thesis of scientific revolutions (1970). One emerging perspective has been proclaimed a 'paradigm shift' (Evans *et al.* 1994a: ix) that will become 'enormously important' and 'destined to transform social and economic policy' (Wilkinson 1996b: 5, ix). It employs the familiar social causation type of explanation, assuming that the roots of social inequalities in health are found in the varying envi-

ronment of social positions. But instead of material shortages and negligent behaviour, focus is now on psychological stress, relative deprivation, and the psycho-social injuries of inequality structures.

To give an overview of this *psycho-social perspective* is the purpose of this paper. The term 'perspective' is chosen because I consider it a set of related approaches rather than a unified theory. It can, however, be identified by three core assumptions: (1) the distribution of psychological stress is an important determinant of health inequalities in present-day affluent societies, (2) psychological stress is strongly influenced by the quality of social and interpersonal relations, and (3) the latter are determined to a large extent by the magnitude of society's inequalities. First, I will make some remarks on the aetiological basis. Then follows a presentation of what I consider the four main sources of the perspective: the *social stress* approach, the related *self-efficacy* approach, the newer *sociology of emotion*, and the *social cohesion* approach. I will highlight elements which I believe are of particular interest for the study of health inequalities. Through this, I also intend to contribute to the construction of this perspective; although there are many controversial issues, I believe it has many promising aspects. In the last section I will however question the 'paradigm shift' thesis: could the psychosocial perspective more properly be described as an enrichment of existing perspectives?

Aetiology

Any explanation of social inequalities in health must combine a general understanding of what causes health and illness with analyses of particular societies. This combination is often what makes the study of health inequalities fruitful, both for medicine and sociology. If current ideas about illness causation seem insufficient to account for observed health inequalities, other aetiological notions are required; attempts to explain health inequalities may on the other hand direct attention to previously neglected aspects of the workings of society.

The psycho-social perspective is a case in point. That grief, loneliness, and similar distressed feelings can impair health, have been part of lay notions of illness for centuries. Since the 1960s a growing number of academic studies have also addressed this topic. Accumulating doubts about the material deprivation explanation for health inequalities has further stimulated interest in the psycho-social environment. The aetiological basis for the psycho-social perspective is the health-damaging potential of psychological stress. Two somewhat different pathways from stress to poor health are proposed: a *direct* effect on disease development, and an *indirect* route, when stress is expressed by health-damaging behaviour.

The indirect pathway implies that people react to adverse circumstances by excessive alcohol use, smoking, accident-prone behaviour, etc., some-

times as conscious or subconscious self-destructive acts, sometimes more 'innocently' in order to alleviate stress. That health can suffer from such behaviour is well documented – and sometimes self-evident, as in the case of violence and homicide.

More controversial is the direct pathway – not as regards mental illness, but as regards somatic disease: can it occur because of psychological stress? Generally, this implies that experiences transmitted by the central nervous system provoke changes in other human organs in a way that threatens health (Kelly *et al.* 1997). Many recent reviews are in favour of this view (see for instance Maes *et al.* 1987, Thoits 1995, Uchino *et al.* 1996, Cohen and Herbert 1996). Evans *et al.* state that 'There is now no longer room for doubt as to the existence of a complex web of linkages, having important implications for health, between the nervous system and other body systems' (1994b: 182). Others are not so sure (see, for instance, Davey Smith and Egger 1996). Despite a rapidly increasing number of studies, reflecting the expansion of fields such as psychoneuroimmunology and psychoneuroendrocrinology, uncertainty remains.

The question is not whether *some* health-related bodily changes may follow from mental appreciations of external circumstances. Recent reviews leave few doubts that psychological stress, generated by despairing circumstances, unsurmountable tasks, or lack of social support, can influence disease-related parameters. Examples are found as regards the cardiovascular system (*e.g.* systolic/diastolic blood pressure), the endocrine system (*e.g.* secretion of catecholamine and cortisol), and the immune system (*e.g.* number of T-cells and Natural Killer cells) (Uchino *et al.* 1996, Cohen and Herbert 1996, Kelly *et al.* 1997, Kiecolt-Glaser and Glaser 1995). The problem is, however, to demonstrate that such changes are large enough and longterm enough to affect health in a significant way. In other words: is the onset of somatic disease, its course, and recovery from it, influenced by psychological stress in a way that makes a difference?

For practical, not to mention ethical, reasons it is difficult to answer this question. A number of observational studies, laboratory experiments, and interventions provide strong indications (see Uchino *et al.* 1996), but most studies utilise only a limited set of variables, and the observed effects of psycho-social factors may sometimes include unmeasured effects of other factors as well. Sufficiently longitudinal approaches are still rare (but see Kelly *et al.* 1997: 438). Laboratory experiments, intervention studies, and animal studies (as regards the latter, see Evans *et al.* 1994b, Wilkinson 1996b; 193pp) have produced suggestive evidence, but under conditions that deviate considerably from 'real' human life. Accordingly, strong conclusions are difficult to draw, not only because of the above-mentioned reasons, but also because diverging findings occur (for instance Carroll *et al.* 1995).

How to conceptualise the role of psychological stress in aetiological processes is moreover a disputed topic. That social inequalities are found

for a wide range of diseases has been interpreted to indicate that stress has a general negative effect on health (Syme and Berkman 1976, Wilkinson 1996b: 71). If so, the mechanisms can be understood in different ways. Stress can be seen as a specific causal factor in the various chains of events which lead to particular diseases (see Uchino *et al.* 1996). Alternatively, stress could influence the body's general capacity for homeostasis, by conditioning 'biological responses . . . in ways that lead to systematic differences in resilience and vulnerability to disease' (Kelly *et al.* 1997: 438). Thus, Kiecolt-Glaser and Glaser (1995) found evidence that cancer, infections, and disorders associated with ageing share the same type of psycho-socially mediated immuno-suppressive mechanisms. Notions of stress-related *general susceptibility* is often part of the aetiological understanding of the psycho-social perspective. Challenges to traditional concepts of disease entities are also raised: Evans suggests that specific diseases should not primarily be understood as endpoints, but rather as alternative pathways from stressful circumstances to illness (1994: 7).

In summary, the aetiological basis for the psycho-social perspective is complex. That psychological stress may contribute to mental illness and health-damaging behaviour is usually agreed; the direct link to specific somatic diseases, or to somatic disease in general, is more contested. One objection to the psycho-social view is that disease aetiology, even in the most affluent societies, cannot be constructed solely in terms of psychological stress. Both the psycho-social *and* the physical environment, and their interaction and interpenetration over time, are perhaps imperative elements of any convincing aetiology, and correspondingly the key to understand health inequalities. I will leave these questions here, and proceed from the 'medical' aspects of the psycho-social perspective to its 'social' aspects: through what kind of processes does psychological stress emerge, and how are they related to society's social inequalities?

The social stress approach

The social stress approach developed especially from Hans Selye's investigations, during the 1940s and 1950s, of physiological responses to external exigencies (see Polloci 1988: 384). In this tradition, stress was defined as 'a state of arousal resulting either from the presence of socioenvironmental demands that tax the ordinary adaptive capacity . . . or from the absence of the means to attain sought-after ends' (Aneshensel 1992: 16). Thus, socioenvironmental demands – stressors – engender psychological stress, *i.e.* a troubled state of the mind which can surface in many ways, as anxiety, fear, hopelessness, or anger.

Stress research is extremely multifaceted (see, for instance, Thoits 1995). Two main tendencies, one more medically, the other more sociologically, oriented, can be distinguished (Aneshensel 1992, Aneshensel *et al.* 1991,

Pearlin 1989). While the former is primarily concerned with consequences of stress in terms of ill health, the latter focuses on the social origins of stress, and includes social inequalities and the stratification system in its models. The two tendencies utilise social variables differently: medically oriented researchers employ information on social background as controls in order to accentuate the links from stressors to stress reactivity, while sociologically oriented researchers focus on how the distribution of stressors depends on people's location within social structures.

The sociologically oriented tendency within stress research can be regarded as the original source of the psycho-social perspective. To explain social inequalities in health has always been part of its agenda. Since the 1960s, various developments have further underlined how people's location in the social structure could result in health differences. What I have in mind are four reorientations: from *any* type of stressors to *negative* stressors; from the 'objective' features of a situation to the 'subjective' impact; the substitution of chronic strain for life events; and the focus on buffers which could reduce the health-damaging potential of stressors.

Stress research started out with the assumption that any kind of event was a potential threat to health (Pearlin *et al.* 1981: 339). This was in line with Selye's preoccupation with reactions to change itself (see Oatley and Jenkins 1992: 72). Thus, not only divorces and deaths, but also marriages and births, could constitute health-damaging experiences, given that the event required overburdening adaption. Poor empirical evidence for this 'neutral' attitude to stressors led to a focus on negative stressors: stressors characterised as unfortunate and unwelcome. As there was little doubt that less privileged social strata tend to face negative circumstances more frequently than those more favourably located in society, the connection between stress processes and the social gradients in health was underscored.

Second, when negative events were highlighted, the subjective impact of events came into focus, because the definition of 'negative' could hardly exclude subjective evaluations. Attention was directed away from the 'objective' characteristics of stressors, towards the way they were appraised by those afflicted (see, for instance, Lazarus 1993). This development put the social environment on the agenda. Appraisals could hardly be seen as isolated individual judgements, but rather as constructed judgements within a social setting. Prevailing norms, types of classifications, and systems of labelling would influence the subjective impact. If events led to social isolation, loss of respect, or other distressing interpersonal relations, the subjective impact would make matters worse. Less exclusionary attitudes towards 'deviants' and more supporting practices towards those who were hit by misfortunes would alleviate the subjective impact. This directed attention to the quality of society's social relations and networks, the cultural representations it favoured, and furthermore towards how such attributes were related to the material differences in society.

Third, findings indicated that the 'deleterious health effects of life changes are of consistently modest magnitude' (Aneshensel 1992: 17). Thus, the one-sided focus upon short-term, abrupt events, as compared with more longstanding stressors – 'enduring problems, conflicts and threats that many people face in their daily lives' (Pearlin 1989: 245) – was criticised. Chronic strain was given a more prominent place in stress research (see, for instance, Turner *et al.* 1995). When stressors were conceptualised as enduring life problems, the association between stress experiences and location in the social structure became more marked, as longstanding economic problems, difficulties related to subordinate positions in the workplace, and the burden of poor neighbourhoods were typical examples.

A fourth reorientation was to focus on vulnerability, *i.e.* why the impact of similar adverse circumstances differed between individuals. One answer was the existence of buffers. Coping resources of diverse types could help people master and overcome adverse circumstances without suffering health setbacks (Maes *et al.* 1987). Social support is the main example, but almost any type of resource connected to personality, social background, education, and financial resources, has been proposed as a moderator of the harmful effects of stressors. Access to buffering factors is related both to the supportiveness of the social environment, and to a person's location in the social structure. Thus, the connection between social position and the circumstances that influence psychological stress was underlined even more.

In this way, the social stress approach explains how social inequalities in health arise. Stressors, particularly long term, chronic stressors, are unevenly distributed in society, basically in line with its structural inequalities. The impact of stressors depends on their subjective appraisal, which follows not only from stressors' factual character, but also from the distribution of buffering resources. Accordingly, psychological stress could be expected to vary with social position and to result in social variations in health. The other approaches outlined below build in many ways on this social stress model, but they also develop particular points and introduce new aspects, thereby contributing to a more comprehensive understanding.

Developing the social stress model: the self-efficacy approach

Social stress research has brought forward a multitude of concepts and hypotheses, addressing topics such as the types of external circumstances that provoke psychological stress, the types of mechanisms that connect circumstances to stress, and the contextual and buffering factors that will influence this process (see, for instance, Maes *et al.* 1987, Pollock 1988, Thoits 1995, Pearlin 1989). In spite of this diversity, it can be argued that stress research has been biased in two respects. It had tended to view people as *passive recipients* of external circumstances, and it has tended to focus solely on *health deterioration*. Moreover, although the emphasis on chronic

strain introduces a more longitudinal view, it could be argued that the social stress approach has often avoided a lifecourse perspective on health.

The *self-efficacy approach* (see Aneshensel 1992: 27–30) can be considered a development of the social stress model which counters all these drawbacks. It addresses the first of these tendencies by pointing to capabilities and power as crucial stress-protecting factors. Humans are conceptualised as acting subjects and not only as being governed by external and structural forces. By emphasising human agency, connections to fundamental theoretical debates in sociology are made. In addition, the self-efficacy approach also opens up for a life-time perspective and for notions about growth in health.

Self-efficacy can be defined as a 'cognitive orientation attributing outcomes such as success and failure to personal attributes, such as ability and effort' (Aneshensel 1992: 27). Mirowsky and Ross (1984) observe that its meaning is virtually synonymous with other frequently employed terms within stress research, such as mastery, internal locus of control, personal control, perceived control of the environment, and instrumentalism. Opposite concepts are fatalism, external locus of control, powerlessness, and learned helplessness. Whatever the term used, the kernel is 'the extent to which people see themselves as being in control of the forces that importantly affect their lives' (Pearlin *et al.* 1981: 339). Self-efficacy is linked to various other phenomena often studied by stress researchers, such as self-esteem, self-concept, social support, and coping style. To emphasise self-efficacy, instead of other related concepts, is therefore somewhat arbitrary. These phenomena are all correlated, and the division of them into separate units, each with a special term, may be artificial, as the sense of support, mastery, self-esteem, and similar traits within an individual is often a tightly connected bundle.

The feeling of mastery, self-efficacy, and being in control, can be assumed a health-promoting factor by itself, as it goes together with a balanced, non-stressed, state of mind. However, it is more common to regard it as a buffer which protects against the damaging effects of adverse external circumstances. Self-efficacy maintains the belief that one can influence one's lot, and this constitutes a defence against feelings of frustration and hopelessness. Thus, self-efficacy makes people less vulnerable to external stressors, and it can be equated with a kind of mental strength which maintains self-esteem and protects against distress.

What generates self-efficacy? The characteristics of the social environment, and social location, seem to be two important factors. Satisfying interpersonal relations will support and enhance self-efficacy. But this is probably also associated with placement in the social structure. Gecas and Schwalbe suggest that the formation of self-esteem is heavily influenced by 'the possibilities that social structures afford for individuals to engage in efficacious action' (1983: 82). Thoits summarises that '[p]erceived control over life circumstances is inversely distributed by social status' (1995: 60).

Thus, the formation of self-efficacy is rooted in the structural features of society, including its system of material inequalities, but also in the supportiveness of the social environment.

The contribution of the self-efficacy approach to the psycho-social perspective is particularly, therefore, that it specifies *mechanisms* involved in the stress process, and that it takes into account the active, participating, actor. It is moreover closely linked to studies which examine the relationship between work organisations and health. Karasek's hypothesis, made credible by many empirical studies, that work-related stress is not dependent on job demands in themselves, but on the combination of a low level of decision latitude with large job demands (see Peterson 1994), points to the significance of having the capability to influence circumstances. Thus, self-efficacy could be importantly involved in the processes whereby health inequalities are formed by hierarchical work organisations.

Furthermore, the notion that health develops over the lifecourse can be viewed as part of the self-efficacy approach. Thus, is parallels the expanding lifecourse perspective on health inequalities (see, for instance, Carroll *et al.* 1996, Davey Smith *et al.* 1997, Wunsch *et al.* 1996). Self-efficacy is commonly seen as a result of a lifetime development. The importance of early life is not only emphasised by biomedically oriented researchers, but also by researchers who focus on psycho-social circumstances. Attachment theory, for instance, argues that inner feelings of security, with important consequences for stress resistance, result from the characteristics of social relations in infancy and childhood (Fonagy 1996). The notion of a life-span development of one's health potential is a prominent part of Antonovsky's concept 'sense of coherence' (1987: 89–127). The principal components of this concept are that the external world is experienced as comprehensible, manageable, and meaningful (1987: 16–19). Given a strong sense of coherence, resistance against deleterious effects of stress, as well as against physical risk factors, is enhanced. Social variations in health are related to the way the sense of coherence is formed. It is no short-time personal characteristic, but develops through life, influenced by placement in the social structure and the availability of resources.

Moreover, Antonovsky's variant of the self-efficacy approach entails a particular view on health evolution. He argues that a sense of coherence can be a salutogenic factor, *i.e.* a contributor to better health. Often, good health is regarded as a kind of 'natural condition', which prevails as long as the environment does not frustrate it; accordingly, poor health is a negative deviation from the 'natural'. In contrast, Antonovsky's writings imply the plasticity of human health, entailing the notion that health may strengthen and grow, as well as deteriorate. It may develop towards more mental and physical strength and towards a longer life, or in the opposite direction. Correspondingly, Wadsworth sees health in terms of a capital analogy: one may 'add interests to' or 'deplete' one's health capital (1996: 159). Such attempts to theorise health are obviously relevant in view of the remarkable

growth in life expectancy during this century, and suggest explanations for variations in life expectancy changes, for instance between social classes or between different countries.

The sociology of emotions: filling black boxes

Emotions are studied in their own right (see, for instance, Collins 1981, Thoits 1989, Kemper 1990, Oatley and Jenkins 1992), but many have also pointed out their relevance for an understanding of health and illness (Freund 1990, James and Gabe 1996, Williams and Bendelow 1996). The structural resemblance between the social stress model and the standard approach of sociologists of emotions is easily recognised. The former addresses the link between social position, stressors, psychological stress, and health outcomes; the latter examines how the social world is experienced in emotionally loaded categories which have bodily correlates: blushing cheeks when feeling shame, muscular tension when angry, throwing up when feeling disgust, etc. Thus, the link between a person's social, subjective, and corporal existence is emphasised by the sociology of emotions. Accordingly, dichotomies such as mind/body, culture/nature, and society/biology are challenged, because such mutually excluding concepts block the understanding of their inner connections (Williams and Bendelow 1996: 28).

In the context of this paper, I will argue that the study of emotions contributes to the psycho-social perspective by addressing certain unclear aspects of the social stress model. Put simply, an awkward problem for the social stress approach is why experiences, which could be regarded as nothing but perceptions of the external world, could possibly be fateful to health. Why aren't they simply mental reflections arousing no more excitement than a dull TV commercial? The general reason is that significant experiences are emotionally loaded. Lots of experiences are irrelevant. Those which have an important meaning also engender emotional responses. Oatley and Jenkins maintain that 'emotions are usually elicited by evaluating events that concern a person's important needs or goals' (1992: 60). As Lazarus puts it: 'there is a world of difference between a non-emotional and an emotional event' (1993: 11). In general, we would expect that emotional experiences have an impact both on mental and physiological processes, while perceptions without this quality would hardly effectuate significant alterations, neither in behaviour nor in physiological functioning.

Psychological stress should accordingly 'be considered part of a larger topic, the emotions' (Lazarus 1993: 10). Stress is a subset of negative emotions, and the study of psychological stress is a special branch within a more general study of emotions. Thus, the question about how emotions emerge encompasses the more specific inquiry into the origins of psychological stress.

Attempted solutions to this question range from social construction approaches at one end of the scale, to positivistic theories at the other end, with (symbolic) interaction theory somewhere in between (Thoits 1989: 319, William and Bendelow 1996: 30). The quality of interpersonal relations is often at the centre of interest. Thus, Burkitt emphasises the relational background to the formation of emotional dispositions (1997); while Lazarus's 'cognitive-motivational-relational theory of emotion' addresses how processes of appraisal attribute meanings to social encounters, accompanied by various types of emotions (1993).

Such approaches underline the social 'nature' of emotions. However, in order to be relevant for the psycho-social perspective, not only a focus on the micro-world of social relations is required, but also a connection to inequality structures at the macro level. The 'positivist' Kemper is more explicit in this respect: he argues that emotions like security, guilt and fear-anxiety arise from social relations formed by dimensions of power and status (Kemper 1979, see also Thoits 1989: 325, Hochschild 1981). Highly relevant, furthermore, is the existential-phenomenological approach of Freund (1988, 1990), who argues that people exhibit 'emotional modes of being' strongly influenced by their positions in social hierarchies, and that these modes become embodied and expressed through various bodily states, including illness.

In order to support the psycho-social perspective, empirical studies should indicate increasing experiences of negative emotions the further 'down' the social ladder people are located. That the frequency of social disadvantages increases in lower social positions is hardly contested, but are they correspondingly associated with emotional responses of a negative character? One alternative hypothesis could be that negative experiences lose their subjective impact when they are common. Based on historical evidence, Lofland (1985) argues that grief will vary according to the degree of infant mortality: the higher the child mortality, the lower the emotional investment in children, so that shorter and less intense grief follows a child's death. Thus, devastating social circumstances could also produce defence mechanisms against negative emotions. On the other hand, one may conjecture that the information supply in modern Western societies implies that people's standards and expectations are usually formed with reference to a national level, maybe even to the international level. People will therefore not compare their fates only with common life situations within their own social milieus, but will refer to wider circumstances. Thus, the idea that disadvantaged social groupings become 'emotionally immune' against their misfortunes because they know of nothing else, is perhaps less viable in present-day Western societies.

To clarify such topics is a challenge for the sociology of emotions, but also of considerable interest for the further development of the psycho-social perspective on health inequalities. It can be added that investigations of the social distribution of negative emotions present many methodological difficulties. Not only are emotions subjective, but they are often partly

subconscious and even prelinguistic. Standard research techniques based on verbal responses may therefore miss the target, and this difficulty is an obstacle to empirical studies of how negative emotions are involved in the generation of health inequalities.

The social cohesion approach

The current interest in psycho-social explanations of social inequalities in health has been spurred on not least by the research of Richard G. Wilkinson (1986, 1990, 1992, 1993, 1994, 1996a, 1996b, 1997). Paradoxically, his interest in psycho-social explanations followed from empirical studies using an indicator of material circumstances, income, as the principal independent variable. His findings, now widely diffused, indicate that in present-day affluent societies, the absolute level of income is no straightforward determinant of health. As absolute deprivation is reduced, health variations become more linked to relative deprivation (or relative income or relative poverty), and life expectancy is influenced by the magnitude of society's income inequalities.

The use of the concept relative deprivation invites some comments. It originated in social psychology around 1940s, as a hypothesis that 'people take the standards of significant others as a basis for self-appraisal and evaluation' (Merton 1967: 40). Contrary to commonsense notions that self-appraisals were linked to absolute standards, it was claimed that people evaluate their situation according to how it compares with others. How people came to choose reference groups, how the processes of social comparisons took place, and what kind of norms of fairness were involved, were central topics of these studies. If people were denied the standards of their reference group, a feeling of deprivation was supposed to follow, *i.e.* disappointment, frustration, and similar distressed feelings because one experienced a standard below what one felt entitled to. This 'classic' relative deprivation model has been applied to the study of mental health (Wagner 1993, Sheeran *et al.* 1995), as well as topics such as migration, absenteeism, criminality, subjective well-being, and especially political attitudes (see Runciman 1966).[1]

This model of relative deprivation, involving reference groups, norms of fairness, and social comparison processes, has however seldom been utilised in its stringent form by researchers of health inequalities. The concept is often introduced without definitions (*cf.* Hasan 1989: 384, Marmot 1994, Wennemo 1993) and used more or less as a shorthand for the view that material standards probably do not damage health directly; nevertheless people's health will vary inversely to their position in society's hierarchical order due to various 'psychosocial processes'.

With Wilkinson's more recent writings (see especially 1996b), a more general framework which addresses the transforming of social inequalities

into health inequalities has been suggested. The kernel is that social inequalities, *i.e.* not only income inequalities, but also power inequalities (for instance, authoritarian hierarchies and non-democratic social organisations) and status inequalities (for instance, as between the two genders, or between ethnic groups), have a fundamental influence on the content of social relations and interactions. The greater the social inequalities (longer distances from top to bottom of the income scale, more authoritarian patterns in families, schools, etc.), the more will the quality of social relations suffer. Inequalities will tend to produce anger, frustration, hostility, fear, insecurity, and other negative emotions. Material inequalities will often go together with fear of, or the actual distressing experience of, failures to secure a socially acceptable material standard of living. Authoritarian power patterns engender feelings of hostility and anger. Differences in status produce contempt from those above and fright and insecurity among those below. Thus, an overall association is assumed to exist between the amount of inequalities in society and the amount of negative feelings and emotions signifying psychological stress. From this, health problems would follow, along both the direct and indirect pathway as described above in the section on aetiology. Smaller social inequalities are, on the other hand, associated with better social relations, *i.e.* more trust, more security, more social support, more self-esteem and self-respect, and more sense of belonging; and also with less financial insecurity and fewer feelings of being materially disadvantaged. Democratic, participatory styles in social organisations, from the family to the political system, ensure self-respect and feelings of being appreciated by one's surroundings, and have therefore additional health-enhancing effects.

Wilkinson substantiates his view both by studies of stress and by accounts of actual societies (1996b: 113–36). He points out, for example, that the low death rate of Roseto (a small US town), which has astonished researchers, could be related to its egalitarian ethos; that variations in infant mortality between Italian regions depend on the degree of 'civic community' which in its turn is associated with the magnitude of social inequalities; and that the remarkable increase in longevity in Japan should be explained not only by economic progress but also by the increased sense of community and the reduced rigidity of the stratificational order developing since 1945.

Social cohesion is accordingly a key concept. It implies a *Durkheimian* view (Durkheim 1964) that 'social facts', *i.e.* societal and collective traits, are more than the sum of individual attributes, and that individuals cannot be understood without grasping the collectivities they are part of. This contrasts with the approaches discussed above, which are mainly individualistic in focus. Wilkinson not only asks questions relating to individual variations in health, but also addresses macro health profiles such as population health and average life expectancy. Instead of analysing individual health as functions of individual variations in income and health behaviour,

Wilkinson advances the importance of global characteristics of social life, for instance in terms of the amount of social capital. This has been defined as 'networks, norms and trust . . . that enable participants to act together more effectively to pursue shared objectives' (Wilkinson 1996b: 221, quoting Putnam 1995: 664–5). The underlying hypothesis is that the degree of social inequalities is highly influential on the formation or deformation of social capital, which is, in its turn, significant for people's health. This theme has also been dealt with by several newer studies, for instance a study of 39 American states (Kawachi *et al.* 1997) which found associations between average mortality, levels of trust, density of group memberships, and income inequality.

Critical views have been raised. There are methodological objections to the studies finding associations between income inequality and population mortality (Judge 1995, but see also Kawachi and Kennedy 1997); and West has maintained that the evidence has 'striking gaps' (1997). The association between population mortality and social inequality is suggestive; nevertheless, this does not release one from examining why the level of individual (or household) income is still closely related to mortality risk within most present-day affluent societies (see, for instance, Davey Smith *et al.* 1996). The hypothesis that material circumstances are without significant influence after certain threshold levels are surpassed, is not necessarily true, and increasing material wealth could have a positive health effect also on higher levels than has often been assumed. Or could it be that the focus on current income is irrelevant, compared with accumulated material disadvantages over the lifecourse (Davey Smith 1996, Bartley *et al.* 1997)? Moreover, Wilkinson's underlying social theory is not always convincing. At the base of the social cohesion approach is the idea that the larger the material inequalities, the more will social environments that lack social support prevail. Thus, it is assumed that social life is more or less directly determined by material structures, which arguably is a much too uncomplicated answer to the perennial question about the relationship between society's material, social, and cultural structures.

In my view, the main contribution of the social cohesion approach to the psycho-social perspective is that it takes components of individual-centred approaches and develops them at the macro level. Wilkinson shows that individual attributes are closely related to collective characteristics. Individual experiences of hostility and social support parallel, on the level of society, the overall occurrence of altruism, trust, and generosity. Individual traits correspond to collective phenomena, and individual processes are insufficiently understood when it is overlooked that individuals are to a large degree formed by the social facts of society, its inequality structures, and its social capital.

Concluding remarks

The psycho-social perspective is a striking and promising attempt to comprehend crucial questions asked by contemporary research regarding, for instance, the persistence of social inequalities in health even in the most affluent societies, the 'challenge of the gradient', and why the growth in life expectancy varies considerably between countries. The social stress approach, especially through its reorientations since the 1960s, provides an overall understanding of why health inequalities mirror social inequalities, in terms of the social distribution of psychological stress. By emphasising the role of human agency, the self-efficacy approach makes up for previous theoretical deficits, and advances the significance of developments over the lifecourse. The sociology of emotions refines the understanding of how external circumstances are transformed into troubling mental states, and the social cohesion approach raises these themes from an individual-centred view to the macro level. Various pieces of the puzzle seem to fall into place when integrating the four approaches, and an encompassing view on health, health inequalities, and its determinants is constructed, based on aetiological notions about the significance of the psycho-social environment and its repercussions in terms of psychological stress. Noteworthy, moreover, is its potential for addressing not only socioeconomic differentials but also health differences related to other dividing lines in society such as gender and ethnicity. In addition, it unites health sciences and social studies to a higher degree than has been usual, since its basic theme is the *social* determination of health and illness.

But does it signal a coming paradigm shift? A paradigm can be viewed as an inclusive set of related propositions which contrasts other understandings and can even replace them, and which is based – at least to a considerable degree – on empirical evidence. Various features of the psycho-social perspective correspond to this: it is comprehensive, consists of several strategies which share, more or less, key concepts, and is informed by numerous empirical studies. Furthermore, it may aim at reinterpreting not only present, but even past health inequalities. A logical next step would be to ask whether psycho-social deprivation has been a major cause all the time, even in the early days of industrialisation, but overlooked by investigators both because material inequality was overwhelming and because natural science has been more prestigious.

Nevertheless, the paradigm metaphor is perhaps too demanding. West (1997) argues that although impressive, the psycho-social perspective does certainly not nullify findings of other perspectives. Many of its suppositions are in need of clarification, for instance the assumption that the distribution of psychological stress reflects, more or less directly, society's inequality structures. The focus on psycho-social environments is certainly justified, as many bits of evidence point to their contribution. But there is a difference

between 'contribute' and 'determine', and current research has not provided evidence that multifactor aetiology, including both psycho-social *and* physical environments, should be succeeded by a one-sided emphasis on psychological stress. Social classes differ also as to their physical milieus, even in the most affluent parts of the world, and the extent to which this has direct repercussions on the body's well-being as well as being mediated by the central nervous system, throughout people's biographies, should be further examined.

Lastly, the claim that the psycho-social perspective will revolutionise our ideas should not hinder us from acknowledging that there are also similarities between the psycho-social perspective and a material deprivation perspective. Both would support Wilkinson's statement that 'the extent of material inequality is a major determinant of population health' (1996b: 9); thus, both would tend to support policies addressing equity questions. They differ as regards the relative weight of the psycho-social and the material pathway from social structure to health inequalities. The most 'extreme' point of view is that psycho-social pathways are decisively most important in present-day societies (see, for instance, Wilkinson 1996b: 4), but there are also less drastic possibilities, allowing for the relevance of psycho-social links without claiming that this implies that direct consequences of material deprivation are eliminated. Thus, one may suggest that the role of the psycho-social perspective will perhaps not be to institute a paradigm shift, but rather to extend and enrich the social causation explanation of health inequalities.

Address for correspondence: Jon Ivar Elstad, NOVA – Norwegian Social Research, Munthes gt. 29, 0260 Oslo, Norway.
email: Jon.I.Elstad@isaf.no

Acknowledgements

The author thanks two anonymous referees for valuable criticism of a previous draft.

Note

1 Another example is the relative deprivation approach to the study of poverty (Townsend 1979), which can be regarded as a more sociological application of the concept. Here, relative deprivation (*i.e.* relative poverty) is considered to be an *objective state*, which occurs when people's level of resources makes them unable to take part in the customs and activities considered normal in their society. Relative poverty implies a withdrawal from normal social life, which occurs when access to resources drops below a critical point. Thus, this concept of relative deprivation does not exclude the feeling of deprivation, but emphasises in

particular ways of life characterised by social isolation and non-participation. It could be added that the relative deprivation theory stands in contrast, in some respects, to the theory of cognitive dissonance (Festinger 1957), which claims that a typical reaction to distressing circumstances is to adjust aspirations in order to reconcile them with actual circumstances. The cognitive dissonance pattern modifies reference points in order to diminish the gap between reality and aspirations, while relative deprivation would occur when expectations are relatively inflexible.

References

Adler, N.E., Boyce, T., Chesney, M.A., Cohen, S., Folkman, S., Kahn, R.I. and Syme, S.L. (1994) Socioeconomic status and health: the challenge of the gradient, *American Psychologist*, 49, 1, 15–24.

Aneshensel, C.S. (1992) Social stress: theory and research, *Annual Review of Sociology*, 18, 15–38.

Aneshensel, C.S., Rutter, C.M. and Lachenbruch, P.A. (1991) Social structure, stress, and mental health: competing conceptual and analytic models, *American Sociological Review*, 56, 166–78.

Antonovsky, A. (1987) *Unraveling the Mystery of Health. How People Manage Stress and Stay Well.* San Francisco: Jossey-Bass Publishers.

Baird, P.A. (1994) The role of genetics in population health. In Evans, R.G., Barer, M.L. and Marmor, T.R. (eds) *Why are Some People Healthy and Others not? The Determinants of Health of Populations?* Berlin/New York: de Gruyter.

Bartley, M., Blane, D. and Montgomery, S. (1997) Socioeconomic determinants of health. Health and the life course: why safety nets matter, *British Medical Journal*, 314, 1194–6.

Blane, D., Davey Smith, G., and Bartley, M. (1993) Social selection: what does it contribute to social class differences in health? *Sociology of Health and Illness*, 15, 1–15.

Blane, D., Brunner, E. and Wilkinson, R.G. (1996) The evolution of public health policy: an anglocentric view of the last fifty years. In Blane, D., Brunner, E. and Wilkinson, R.G. (eds) *Health and Social Organization. Towards a Health Policy for the 21st Century.* London and New York: Routledge.

Burkitt, I. (1997) Social relationships and emotions, *Sociology*, 31, 37–55.

Carroll, D., Davey Smith, G., Sheffield, D., Shipley, M.J. and Marmot, M.G. (1995) Pressor reactions to psychological stress and prediction of future blood pressure: data from Whitehall II study, *British Medical Journal*, 310, 771–6.

Carroll, D., Davey Smith, G. and Bennett, P. (1996) Some observations on health and socio-economic status, *Journal of Health Psychology*, 1, 1, 23–39.

Cohen, S. and Herbert, T.B. (1996) Health psychology – psychological factors and physical disease from the perspective of human psychoneuroimmunology, *Annual Review of Psychology*, 47, 113–42.

Collins, R. (1981) On the microfoundations of macrosociology, *American Journal of Sociology*, 86, 984–1014.

Dahl, E. and Kjærsgaard, P. (1993) Trends in socioeconomic mortality differentials in post-war Norway – evidence and interpretations, *Sociology of Health and Illness*, 15, 447–71.

Davey Smith, G. (1996) Income inequality and mortality: why are they related? *British Medical Journal*, 312, 987–8.

Davey Smith, G., Neaton, J.D., Wentworth, D., Stamler, R. and Stamler, J. (1996) Socioeconomic differentials in mortality risk among men screened for the Multiple Risk Factor Intervention Trial: I. white men, *American Journal of Public Health*, 86, 4, 486–96.

Davey Smith, G. and Egger, M. (1996) Commentary: understanding it all – health, meta-theories, and mortality trends, *British Medical Journal*, 313, 1584–5.

Davey Smith, G., Hart, C., Blane, D., Gillis, C. and Hawthorne, V. (1997) Lifetime socioeconomic position and mortality – prospective observational study, *British Medical Journal*, 314, 547–52.

Durkheim, E. (1964 [1895]) *The Rules of Sociological Method.* New York: Free Press.

Evans, R.G. (1994) Introduction. In Evans, R.G., Barer, M.L. and Marmor, T.R. (eds) *Why are Some People Healthy and Others not? The Determinants of Health of Populations?* Berlin/New York: de Gruyter.

Evans, R.G., Barer, M.L. and Marmor, T.R. (1994a) Preface. In Evans, R.G., Barer, M.L. and Marmor, T.R. (eds) *Why are Some People Healthy and Others not? The Determinants of Health of Populations?* Berlin/New York: de Gruyter, ix–xix.

Evans, R.G., Hodge, M. and Pless, I.B. (1994b) If not genetics, then what? Biological pathways and population health. In Evans, R.G., Barer, M.L. and Marmor, T.R. (eds) *Why are Some People Healthy and Others not? The Determinants of Health of Populations?* Berlin/New York: de Gruyter, 161–88.

Evans, R.G. and Stoddart, G.L. (1994) Producing health, consuming health care. In Evans, R.G., Barer, M.L. and Marmor, T.R. (eds) *Why are Some People Healthy and Others not? The Determinants of Health of Populations?* Berlin/New York: de Gruyter.

Festinger, L. (1957) *A Theory of Cognitive Dissonance.* Evanston, Ill.: Row, Peterson.

Fonagy, P. (1996) Patterns of attachment, interpersonal relationships and health. In Blane, D., Brunner, E. and Wilkinson, R.G. (eds) *Health and Social Organization. Towards a Health Policy for the 21st Century.* London and New York: Routledge.

Freund, P.E.S. (1988) Bringing society into the body: understanding socialized human nature, *Theory and Society*, 17, 839–64.

Freund, P.E.S. (1990) The expressive body: a common ground for the sociology of emotions and health and illness, *Sociology of Health and Illness*, 12, 4, 452–77.

Gecas, V. and Schwalbe, M.L. (1983) Beyond the looking-glass self: social structure and efficacy-base self-esteem, *Social Psychology Quarterly*, 46, 77–88.

Hasan, J. (1989) Ways-of-life, stress, and differences in morbidity between social classes. In Fox, J. (ed) *Health Inequalities in European Countries.* Aldershot: Gower.

Hochschild, A.R. (1981) Power, status and emotion, *Contemporary Sociology*, 10, 1, 73–7.

James, V. and Gabe, J. (1996) Introduction: connecting emotions and health? In James, G. and Gabe, J. (eds) *Health and the Sociology of Emotions.* Oxford: Blackwell Publishers, Sociology of Health and Illness Monograph Series.

Judge, K. (1995) Income distribution and life expectancy. A critical appraisal, *British Medical Journal*, 311, 1282–5.

Kadushin, C. (1964) Social class and the experience of ill health, *Sociological Inquiry*, 35, 67–80.

Kaplan, J.R., Pamuk, E., Lynch, J.W., Cohen, R.D. and Balfour, J.L. (1996) Income equality and mortality in the United States, *British Medical Journal*, 312, 999–1003.

Kawachi, I. and Kennedy, B.P. (1997) The relationship of income inequality to mortality: does the choice of indicator matter? *Social Science and Medicine*, 45, 7, 1121–7.

Kawachi, I., Kennedy, B.P. and Lochner, K. (1997) Social capital, income inequality, and mortality, *American Journal of Public Health*, 87, 9, 1491–8.

Kelly, S., Hertzman, C. and Daniels, M. (1997) Searching for the biological pathways between stress and health, *Annual Review of Public Health*, 18, 437–62.

Kennedy, B.P., Kawachi, I. and Prothrow-Stith, D. (1996) Income distribution and mortality: cross sectional ecological study of the Robin Hood Index in the United States, *British Medical Journal*, 312, 1004–7.

Kemper, T.D. (1979) *A Social Interactional Theory of Emotions*. New York: Wiley.

Kemper, T.D. (ed) (1990) *Research Agendas in the Sociology of Emotions*. New York: State University of New York Press.

Kiecolt-Glaser, J.K. and Glaser, R. (1995) Psychoneuroimmunology and health consequences: data and shared mechanisms, *Psychosomatic Medicine*, 57, 269–74.

Kuhn, T. (1970) *The Structure of Scientific Revolutions*. 2nd Edition. Chicago: University of Chicago Press.

Lazarus, R.S. (1993) From psychological stress to the emotions: a history of changing outlooks, *Annual Review of Psychology*, 44, 1-21.

Link, B.G. and Phelan, J. (1995) Social conditions as fundamental causes of disease, *Journal of Health and Social Behavior*, Extra Issue, 80–94.

Lofland, L.H. (1985) The social shaping of emotion: grief in historical perspective, *Symbolic Interaction*, 8, 171–90.

Maes, S., Vingerhoets, A. and Vanheck, G. (1987) The study of stress and disease – some developments and requirements, *Social Science and Medicine*, 25, 6, 567–78.

Marmot, M.G. (1986) Social inequalities in mortality: the social environment. In Wilkinson, R.G. (ed) *Class and Health. Research and Longitudinal Data*. London and New York: Tavistock Publications.

Marmot, M.G. (1994) Social differentials in health within and between populations, *Dædalus*, 123, 4, 197–216.

Merton, R.K. (1967) *On Theoretical Sociology. Five Essays, Old and New*. New York and London: The Free Press.

Mirowsky, J. and Ross, C.E. (1984) Mexican culture and its emotional contradictions, *Journal of Health and Social Behavior*, 25, 1, 2–13.

Oatley, K. and Jenkins, J.M. (1992) Human emotions: function and dysfunction, *Annual Review of Psychology*, 43, 55–85.

Pearlin, L.I. (1989) The sociological study of stress, *Journal of Health and Social Behavior*, 30, 3, 241–56.

Pearlin, L.I., Menaghan, E.G., Lieberman, M.A. and Mullan, J.T. (1981) The stress process, *Journal of Health and Social Behavior*, 22, 337–56.

Peterson, C.L. (1994) Work factors and stress – a critical review, *International Journal of Health Services*, 24, 3, 495–519.

Pollock, K. (1988) On the nature of social stress – production of a modern mythology, *Social Science and Medicine*, 26, 3, 381–92.

Putnam, R.D. (1995) Tuning in, tuning out: the strange disappearance of social capital in America, *Political Science and Politics*, December, 664–83.

Runciman, W.G. (1966) *Relative Deprivation and Social Justice. A Study of Attitudes to Social Inequality in Twentieth-century England.* London: Routledge and Kegan Paul.

Sheeran, P., Abrams, D. and Orbell, S. (1995) Unemployment, self-esteem, and depression – a social-comparison theory approach, *Basic and Applied Social Psychology*, 17, 1–2, 65–82.

Syme, S.L. and Berkman, L.F. (1976) Social class, susceptibility and sickness, *American Journal of Epidemiology*, 104, 1, 1–8.

Tarlov, A.R. (1996) Social determinants of health: the sociobiological translation. In Blane, D., Brunner, E. and Wilkinson, R.G. (eds) *Health and Social Organization. Towards a Health Policy for the 21st Century.* London and New York: Routledge.

Thoits, P.A. (1989) The sociology of emotions, *Annual Review of Sociology*, 15, 317–42.

Thoits, P.A. (1995) Stress, coping, and social support processes – where are we – what next, *Journal of Health and Social Behavior*, Extra Issue, 53–79.

Townsend, P. (1979) *Poverty in the United Kingdom.* Harmondsworth: Penguin.

Turner, R.J., Wheaton, B. and Lloyd, D.A. (1995) The epidemiology of social stress, *American Sociological Review*, 60, 104–25.

Uchino, B.N., Cacioppo, J.T. and Kiecolt-Glaser, J.K. (1996) The relationship between social support and physiological processes: a review with emphasis on underlying mechanisms and implications for health, *Psychological Bulletin*, 119, 3, 488–531.

Vågerö, D. and Illsley, R. (1995) Explaining health inequalities: beyond Black and Barker. A discussion of some issues emerging in the decade following the Black Report, *European Sociological Review*, 11, 3, 219–41.

Wadsworth, M. (1996) Family and education as determinants of health. In Blane, D., Brunner, E. and Wilkinson, R.G. (eds) *Health and Social Organization. Towards a Health Policy for the 21st Century.* London and New York: Routledge.

Wagner, R.M. (1993) Psychosocial adjustments during the 1st year of single parenthood – a comparison of Mexican-American and Anglo women, *Journal of Divorce and Remarriage*, 19, 1–2, 121–42.

Waldmann, R.J. (1992) Income distribution and infant mortality, *Quarterly Journal of Economics*, 107, 1283–302.

Wennemo, I. (1993) Infant mortality, public policy and inequality – a comparison of 18 industrialized countries 1950–85, *Sociology of Health and Illness*, 15, 4, 429–46.

West, P. (1997) (Book Review) Wilkinson, R.G., Unhealthy societies: the afflictions of inequality, *Sociology of Health and Illness*, 19, 5, 668–70.

Wilkinson, R.G. (1986) Income and mortality. In Wilkinson, R.G. (ed) *Class and Health. Research and Longitudinal Data.* London and New York: Tavistock Publications.

Wilkinson, R.G. (1990) Income distribution and mortality: a natural experiment, *Sociology of Health and Illness*, 12, 4, 391–412.

Wilkinson, R.G. (1992) Income distribution and life expectancy, *British Medical Journal*, 308, 1113–14.

Wilkinson, R.G. (1993) The impact of income inequality on life expectancy. In

Platt, Stephen, Thomas, Hilary, Scott, Sue and Williams, Gareth (eds) *Locating Health. Sociological and Historical Explorations*. Aldershot: Avebury.

Wilkinson, R.G. (1994) The epidemiological transition: from material scarcity to social disadvantage, *Dædalus*, 123, 4, 61–77.

Wilkinson, R.G. (1996a) How can secular improvements in life expectancy be explained? In Blane, D., Brunner, E. and Wilkinson, R. (eds) *Health and Social Organization. Towards a Health Policy for the 21st Century*. London and New York: Routledge.

Wilkinson, R.G. (1996b) *Unhealthy Societies. The Afflictions of Inequality*. London and New York: Routledge.

Wilkinson, R.G. (1997) Health inequalities: relative or absolute material standards? *British Medical Journal*, 314, 22, 591–8.

Williams, S.J. and Bendelow, G. (1996) Emotions, health and illness: the 'missing link' in medical sociology?: In James, V. and Gabe, J. (eds) *Health and Sociology of Emotions*. Oxford: Blackwell Publishers, Sociology of Health and Illness Monograph Series.

Wunsch, G., Duchêne, J., Thiltgès, E. and Salhi, M. (1996) Socio-economic differences in mortality. A life course approach, *European Journal of Population*, 12, 167–85.

Paper III:

"Inequalities in health related to women's marital, parental, and employment status – a comparison between the early 70s and the late 80s, Norway", *Social Science & Medicine*, 42 (1): 75-89, 1996.

Pergamon

0277-9536(95)00078-X

Soc. Sci. Med. Vol. 42, No. 1, pp. 75–89, 1996

0277-9536/96 $15.00 + 0.00

INEQUALITIES IN HEALTH RELATED TO WOMEN'S MARITAL, PARENTAL, AND EMPLOYMENT STATUS—A COMPARISON BETWEEN THE EARLY 70s AND THE LATE 80s, NORWAY

JON IVAR ELSTAD

Institute of Applied Social Research, Munthesgt. 31, 0260 Oslo, Norway

Abstract—Studies indicate that inequalities in women's health are associated with women's marital, parental and employment status. The causal mechanisms which generate these inequalities are linked to social change at the macro level. The present study asks whether patterns of ill-health according to women's statuses have changed during recent decades in Norway. Five national surveys 1968–91 are analyzed, using number of long-standing diseases as an indicator of health. The results indicate that health differences between full-time employed women and other employment statuses have increased during the 70s and 80s. As regards marital and parental status, the observed changes are not significant. The findings suggest that important causal mechanisms generating health differences related to marital status are located in the private sphere. The interpretation of the widening health gap between employed and non-employed women focusses both on developments which have made it easier to combine employment and family duties, on new norms which favour the self-esteem of employed women, and on health selection processes connected to welfare state developments.

Key words—health inequalities, women, marital, parental, employment status, long-standing diseases, trend

INTRODUCTION

Inequalities in women's health are associated with the statuses women occupy in family and work. A number of studies, both in the U.S. [1–14], the U.K. [15–21], and in Norway [22, 23], indicate that married women usually have better health than previously married women (divorcees and widows), that mothers are usually healthier than childless women, and that women who are employed outside the home usually experience less ill-health than women without such employment.

Thus, women's health is linked to their marital, parental and employment status. Research on these connections has taken several directions. It has been asked whether the associations are equally valid for all aspects of health, for example physical health and mental health [1, 19]; whether they are similar in different socio-economic layers of society [15, 16]; and whether combinations of these three statuses produce any interactional effects [4, 9, 14]. Considerable efforts have also been spent on trying to discover the causal mechanisms behind these relationships [2, 5, 9–13, 19, 20, 24–31].

One important question concerning these associations has been neglected, however. This is the question about trends over time. Are the associations between women's health and their marital, parental and employment status stable, or are they changing? And if changing, in what direction?

The study of Verbrugge and Madans [11] addressed this question, based on data from the U.S. National Health Interview Survey 1964–65 and 1977–78. In the main, they found "stability of role effects", but also that differences had become "clearer and more distinct"—in other words, a weak tendency to more differentiation related to women's statuses. Apart from this study, it is hard to find research on the question. The main reason for this neglect is probably that studies of trends need surveys or other types of data which collect comparable health indicators and information about women's marital, parental end employment status over a prolonged time period. Such data sets are rare, and if they exist, they have not been employed for this purpose.

The present article will examine the linkages between health and women's marital, parental and employment status in Norway during recent decades. First, I will discuss, in general terms, various processes which can link women's statuses to their health. After that, I will outline some reasons for hypothesizing that these linkages have changed in Norway during the 70s and 80s, before I turn to the analysis of data from five Norwegian national surveys, conducted in the period 1968–1991.

A general perspective

In order to understand why associations between women's statuses and their health may change over

time, one must consider both the *causal mechanisms* which generate associations between women's statuses and their health, and the *macro-social level*, which influences the operation of these causal mechanisms.

Causal mechanisms. By causal mechanisms I refer to the immediate causal processes which have the capability of generating inequalities in health associated with women's family and employment statuses. Most common are 'social causation' models, which assume that women's health is a consequence of their employment status, and 'health selection' models (also labelled 'social selection' models, *cf* [32]), which assume that the causal direction is the other way round. Other explanatory models have also been discussed, for instance the 'artefactual model' [33] and the 'indirect selection model' [34–36], but in this paper, I limit my discussion to the two main models.

Social causation models search for intervening variables linking women's marital, parental, or employment status to women's health. The idea is that various factors are connected to each status, and each of these factors has either beneficial or detrimental effects. Negative effects connected to one status may be neutralized or overshadowed by positive effects of the same status, or by positive effects of other statuses, and the resulting health consequences depend on the sum-total of such advantageous and harmful effects [14, 17, 19].

Often, the aetiological reasoning of social causation models emphasizes psychological factors. It is assumed that "social and environmental factors influence a person's basic feelings about life (negative and positive), which in turn influence physiological processes that lead toward or protect from physical and mental illness" [8] (p. 237); also [17] (p. 376). Commonly, it is asked whether women's various statuses provide social support [5, 7], self-esteem [13, 31], self-realization [37] and personal control [29, 30, 38].

An alternative social causation model focusses on resources and material circumstances of various kinds, such as education, material standard of living and working conditions. The question has been raised whether the health effects of women's employment depend on type of work [8, 12], on domestic workload [19], or on the combination of working conditions at home and in paid employment outside the home [20].

The common characteristic of *health selection explanations* is the assumption that the associations between health and social positions are generated by social processes which, in some way or another, "select" persons because of their health [32, 34–40]. Many studies acknowledge the possibility that health may influence women's marital, parental and employment status [9–12, 15–17, 19–29], and this causal mechanism is highly relevant in the study of inequalities in women's health. In spite of this, studies which examine the health selection mechanism in this field are rare [16] (p. 426), probably because one seldom finds the data required for this type of research.

Macro-social level and social change. The operation of causal mechanisms, however, is linked to the macro-social context. Causal mechanisms do not function independently of society—rather, their workings are influenced by political, economic and cultural characteristics at the macro level. For instance, the connection between paid employment and women's health depends on the typical working conditions experienced by the female work force, which in turn depend on the technological level of production, the gendered division of the labour market, and laws and rules which regulate the work-places. Another example: a number of institutional arrangements influence the scope of health-related selection: demanding production norms may increase health related discrimination in the labour market, governmental efforts may facilitate the employment of chronically ill persons, and disability pensions may boost exits from the labour market.

An awareness of the interconnections between the various causal mechanisms and the macro-social level is often absent in studies on this theme. Notable exceptions are the research of Arber and Lahelma [18, 25], who focus on circumstances such as prevailing norms, welfare state institutions, characteristics of the labour market and gender relations, when analyzing inequalities in women's health.

Because of these interconnections between the causal mechanisms and the political, economic and cultural characteristics of society, the associations between women's statuses and their health may change through time. The social context determines which causal mechanisms are important, and, as a consequence of social change, the operation of the causal mechanisms may be altered. Some mechanisms may be elevated to prominence, while others may be degraded to insignificance. By such developments, the linkages between women's statuses and their health may be generated in a new way, and the result may be new associations between health and women's statuses.

Social change in Norway and some possible consequences

The preceding section has outlined a general perspective which underlines the possibility of changes in the relationships between women's health and their family and employment status. In this section the question is posed in a more specific way: Have there occurred social changes in Norway during recent decades which make it probable that the associations between women's health and their statuses have been altered?

Women's lives have been transformed in many ways in the 70s and 80s. Here, it is of course not possible to render a complete account of these transformations. The purpose of this section is to give an overview of some important developments and to formulate tentative hypotheses about their effects: is it reasonable to expect smaller or larger differences in

health related to women's statuses from these developments? Or is it most probable that the associations between health and status have not been affected at all?

Marital status. Health selection has been discussed as a possible reason for the health disadvantages of previously married women, i.e. divorcees and widows, compared to married women [41], but the social causation model is also regarded as highly relevant [42]. To lose a partner, whether by marital break-up or by death, causes a feeling of loss, and often grief and depression. Especially in the case of divorce, one's self-esteem is often suffering. These psychic problems, which initially are connected to the event of separation from one's partner, may continue for a long time after the break-up. As life without a partner goes on, new problems may emerge: Social support and the satisfaction from close personal relationships are absent, living alone may lead to social isolation and economic difficulties may arise if the partner used to be the main bread-winner. The status of divorced women can be low, and they may experience covert or open hostility from their surroundings. If they are mothers they may fear that the situation is particularly stressing for their children.

Thus, within the social causation perspective, the health disadvantages of previously married women, as compared to married women, are explained by a combination of psychic stress and economic difficulties. Some of the problems faced by divorced and widowed women are also experienced by single, never-married women, but apparently without such harmful consequences—the health state of never-married women is usually quite similar to that of married women [15, 25] (Table 2).

Has social change in Norway eased the situation for previously married women, compared to married women?

Several developments make is possible that the circumstances of previously married women, especially divorced women, have improved. The increase in divorces, from five per 1000 marriages in the early 70s to 11 in the early 90s [43] (p. 292), has resulted in a markedly higher proportion of women who are currently separated or divorced (*cf* Table 1). Experiences of marital break-ups have become more common, and we would expect that this would alleviate the psychic stress: Divorced women are not an "outcast" minority, and divorcees always find many others who share their situation. Formerly, the majority public opinion condemned divorce, but liberal attitudes are now much more widespread. Moreover, many divorced women and some widows remarry, which means, among other things, that previously married women need not be isolated, nor feel that their situation blocks the forming of new, close, personal relationships. Lastly, although it is clear that some divorced women, especially if they also have maternal responsibilities, encounter economic difficulties, there have been some developments which perhaps have improved their material situation: more possibilities for female employment create possibilities for income also among previously married women, more regulations of the obligations of divorced fathers have perhaps assured that they pay for their children to a greater degree than earlier, and several public benefits aimed at securing the material existence of single mothers have been implemented: the number of single parents (mostly women) who received benefits from the National Insurance Fund was $<10{,}000$ in 1970, but around 44,000 in 1992 [44] (p. 128), [45] (p. 211).

This discussion leads to the tentative hypothesis that inequalities related to women's marital status have gradually become smaller, because the disadvantages, both as regards psychic and material stress, among previously married as compared to married women, may have diminished.

Parental status. Usually, health differentials related to women's parental status are small. One study [2] concludes that "(h)aving children in the household generally contributes to poor mental health", but most studies indicate the opposite result that mothers tend to be of better health than women without children [9, 16, 20]. Health selection probably generates part of these inequalities, both because health problems may make it difficult to become pregnant and because women may themselves restrict their number of childbirths because of their health. Also social causation processes may be involved. Childlessness may be a cause of depression [46]. Children are a source of happiness and fulfillment, they provide a sense of meaning to life, they are a source of social support and even a source of self-esteem and prestige, and through their children women may establish wider social relations.

But children may also become a source of worry and disappointment to their mothers (but surely not as often as husbands disappoint their wives), and "(c)hildren are clearly often a source of considerable work and can place all sorts of constraints on time and autonomy" [20] (p. 661). This aspect is perhaps especially relevant when maternal duties are combined with paid employment, and when material resources are scarce. One study based on U.K. data from the middle of the 70s found adverse health consequences among women with children who worked in low-status occupations [17].

In Norway during recent decades, the proportion of women having at least one child has not changed much, but the total fertility (estimated childbirths per woman during her whole childbearing period) has decreased from 2.7 in the late 60s to 1.8 in the late 80s [47] (p. 60). Fewer women have many children (*cf* Table 1), and stress and exhaustion among mothers because they have many children have probably been reduced. The general rise in standard of living—indisputable in the 70s but more modest in the 80s—and expansion of public benefits to mothers and families who have children, have probably made it easier to

provide for children's upbringing in material terms. Thus, the positive aspects of having children may have become more unambiguous, and this may have increased the differences in health between childless women and mothers.

Changes which may affect the health selection mechanism could also point in the same direction. Birth control technology, such as the pill which was introduced in Norway during the 60s, and the implementation of women's right to decide themselves about abortion in 1978, have augmented women's control over their childbirths and at the same time made it possible to abstain from having children because of health difficulties. Thus, it is possible that women with health problems have no or only one child to a greater degree than earlier. But the opposite consequence is also imaginable: it has been argued that the availability of contraception methods has made it possible also for healthy women to avoid pregnancies [11]. If this is the case, then childlessness will be more widespread among women in good health, and the health differentials related to number of children would diminish.

Thus, some social changes can be believed to have increased the health differences between mothers and childless women, but others point in the opposite direction. Our tentative hypothesis is that the association between health and parental status has remained more or less unchanged.

Employment status. Through stressful and dangerous working conditions [8, 37], employment may be detrimental to women's health. In spite of this, studies indicate that employed women have better health than their non-employed 'sisters'. Social causation processes are probably important. Employment may provide social support from one's colleagues, opportunities for self-actualization and a feeling of power and self-confidence from having a job and an income [7, 8, 19, 30]. The relative benefits of employment may arise because the most common alternative, full-time housework, often is monotonous and isolated [2]. The alleviating effect of employment in the case of a stressful family life has also been underlined [7].

But selection processes are probably also involved. Recruitment into employment may be related to health, as women with health difficulties may decide to abstain from paid employment or because employers may discriminate against them. Exits from the labour markets are perhaps even more linked to health: ill women may quit working, or are pressurized into retirement, and thus only healthy women remain in employment (the "healthy worker effect" [39]).

The hypothesis that employment may be harmful to health has often been discussed in terms of "role overload" or "role strain" [17, 19]. Multiple roles, usually examined in the form of combinations of paid work with family obligations, have been hypothesized to lead to physical exhaustion, role conflicts, feelings of insufficiency and other types of psychic stress. Common sense tells us that role overload is felt by many employed women, but studies usually fail to reveal statistically significant harmful effects [4, 9, 14, 19].

The great transformations as regards female work during the 70s and 80s may have had a number of effects as regards the association between employment status and health. In contrast to some decades ago, prevailing norms now support women's employment. In 1993, about two thirds of a representative sample of Norwegian women answered to a survey question that they felt that they were expected to have paid employment [48]. The number of full-time homemakers have dropped dramatically (*cf* Table 1) and their prestige has probably also dropped. Homemaking is often not considered as 'real work', at least when there are no small children in the household. Accordingly, employed women, especially full-time employees, receive more respect from their surroundings, relative to what other employment statuses receive, during the late 80s than previously, and this may have ramifications as regards psychic well-being and health.

Developments which have alleviated strain and stress connected to female employment may also have

Table 1. Distribution of marital, parental and employment status in the samples (%). Women 31–60 years of age, 1968–91

		1968	1975	1985	1987	1991
Marital status	Never married	9.1	7.1	5.2	6.2	5.1
	Married[a]	83.4	83.0	84.7	83.0	85.3
	Divorced[b]	3.2	4.7	6.3	7.8	8.3
	Widow	4.4	5.2	3.8	3.1	1.3
Children under 0–16 years in the household	None	43.6	46.7	43.4	48.6	47.6
	One	22.4	19.5	23.6	20.4	20.0
	Two	17.5	19.5	22.4	22.2	23.2
	Three or more	16.5	14.3	10.6	8.8	9.2
Labour market	Full-time job	27.1	30.9	39.0	45.3	48.2
	Part-time	}	25.9	32.1	30.4	26.1
	Homemaker	} 72.9	38.3	18.6	14.9	12.3
	Others	}	4.8	10.3	9.4	13.4
(*N*)		(2082)	(1981)	(1902)	(975)	(904)

[a]1985–1991, Married and cohabitating.
[b]Divorced or separated.

led to increased differences in health between employed and non-employed women. The standard working week settled by the nation-wide agreements between trade unions and employers has been reduced from 42.5 hr in the early 70s to 37.5 hr in the late 80s, and this could mean that the risk of getting exhausted among women who work a full week has been reduced. Moreover, it is possible that typical female workplaces are less dangerous to health during the late 80s, as compared to two decades earlier. The number of manual factory jobs for women has been reduced, while women's employment in education, the health services and other service occupations has expanded. More women have higher education, and there is a slightly increased percentage of female employees in managerial positions [43] (p. 183, 131). Among employed women, fewer are exposed to the typical risk factors of traditional factory work (pollution, noise, long hours with heavy physical work), and a slowly growing proportion of working women have more qualified jobs, which may benefit health because of the potential for self-realization and personal power (but perhaps career stress is increasing?).

Not only is it likely that the burden of employment is somewhat reduced: also housework has decreased. The average woman spent 4.1 hr per day on housework in 1971/72 and 2.2 hr in 1990/91 [47], (p. 114). Women's total working time in paid employment outside the home and work within the household has been reduced from 8.1 to 7.4 hr 1970–1990 [49] (p. 74), and a parallel reduction has occurred among full-time employed women. Thus, the combination of full-time employment and household tasks may have become easier for women, at least in terms of total working time. Furthermore, the chance that working women experience problematic conflicts when they have children is probably reduced. By the late 80s, there was more public care of children, in terms of longer hours at school and increased availability of daycare for pre-school children. All in all, this means that the harmful potential of paid employment among women probably has been reduced, and the subsequent outcome is likely to be larger health differentials between employed and non-employed women.

As regards selection into paid employment, it is difficult to assess how the dramatical augmention of women's employment could have affected possible selection processes. If health plays an important role when women enter paid employment, we would expect that inequalities between employed and non-employed women have decreased during recent decades. This follows from the general tendency that when a narrowly selected category grows from a small to a large minority, the difference between the selected and the remaining category is diminished (at least when the selection criterion is normally distributed).

However, it is quite doubtful that women during the 60s and 70s were selected into paid employment because of their health. At that time, there was little social pressure that women should work, and no special reason why women in good health should seek paid employment instead of doing housework. This has changed: during the 80s most women will seek paid employment, and prevailing norms prescribe paid employment among women, at least if there are no strong reasons, such as responsibilities for children, that they should not. Perhaps the selective mechanisms have become stronger. The market for female workers shrank during the economic recessions of the 80s, and since women workers were abundant, employers had great opportunity to reject job seekers with health problems.

Another important health selection mechanism is probably the development of welfare state benefits. The percentage of women 16–67 receiving disability pension has doubled during the last two decades (5.1% in 1970, 9.5 in 1992), and also other types of welfare state benefits have increased in scope [44] (p. 109, 128); [45] (p. 109, 211). This has created a new 'employment' status for women, as recipients of disability pension or of other welfare state benefits. Women are very often recruited into this role as a result of bad health. Earlier, working women with health problems may have tried to hang on to work, or they have quit paid employment and become housewives. The creation of these welfare state provisions may have weeded out women with poor health from the work force, and the opportunities to do this have increased during the last decades, as is proved by the growth in the number of women receiving such benefits. In consequence, fewer women with health problems remain in employment, and the difference in health will probably increase.

Summing up. The overview presented above illustrates the complexity of the processes which generate inequalities in health related to women's marital, parental and employment status. On the one hand, numerous aetiological processes and causal mechanisms are involved. On the other hand, these aetiological processes and causal mechanisms are linked to the social context in multiple ways, and social change in a broad sense will have implications for the processes which generate these inequalities.

Because so many factors are involved, it is of course difficult to foresee the net consequences as regards health inequalities among women. Nevertheless, I suggest three hypotheses:

1. Disparities related to marital status have diminished, because the social situation of previously married women, in particular divorced women, has improved.
2. Disparities related to parental status have been stable.
3. Health differentials according to employment status have increased, for several reasons: the disadvantages of female employment in combination with household chores have been alleviated, the social esteem of homemaking has become relatively lower, and selection mechanisms have become more

important due to health related selection into paid employment and to developments of welfare state benefits which have made it easier for ill women to retreat from paid work.

The following analyses will try to shed light on these hypotheses.

DATA AND VARIABLES

The surveys

The Central Bureau of Statistics of Norway has conducted national surveys since the late 60s, and data covering more than 20 years are available. The Health Surveys of 1968, 1975 and 1985 are especially oriented towards questions about health [50–52]. The Surveys of Level of Living 1987 and 1991 [53, 54] provide information which is comparable to the Health Surveys. For the purpose of this paper, these five national surveys are analyzed.

The Health Surveys sampled households and included persons of all ages, while the Surveys of Level of Living sampled individuals aged 16–79. Data were collected by means of personal interviews, or by telephone interviews if visits were rejected. The response rate was not reported in the 1968 data; it was 89% in 1975 and dropped gradually in the succeeding surveys to 75% in 1991. The Central Bureau of Statistics has compared the samples of each survey with the corresponding population register, and these comparisons have revealed no serious bias as regards demographic composition [50–54].

Classification of marital, parental and employment status

The samples analyzed in this paper are described in Table 1. Women aged 31–60 are studied. Non-marital cohabitation, which became quite widespread during the 70s and 80s, and which in many respects is comparable to marriage [55], has been merged with the married category in the 1985–1991 surveys—in the 1968 and 1975 surveys, information about non-marital cohabitation was not collected. It should be noted that remarriage after a divorce is quite common, and that some respondents classified as married have previously been divorced or widowed.

The classification of parental status is based on information about number of household members aged 0–16. Older children, or children who have moved out of the household (who, for instance, live with their father in case of divorce), are not counted. In some cases children in the household are not the offspring of the respondent—they are step-children or children belonging to other members of the household. The variable does not measure parental status in terms of number of childbirths, but rather in terms of daily parental responsibilities as indicated by the number of children living in the household.

The 1975–1991 data have the information necessary for a classification into full-time work (defined as 30 hr or more per week), part-time work (1–29 hr per week), full-time homemaker, and a residual category, termed 'others', which consists mainly of women whose main source of income is disability pension or other types of transfers from the welfare state. While homemakers and employed outside the home covered fairly well the main employment statuses of women in the 60s, this dichotomy became less satisfying during the 70s and 80s, when the number of non-employed women with income from the welfare state increased considerably. The 1968 survey, however, asked only whether or not paid work was the main activity or main source of income. It seems legitimate to suppose that paid employment which is one's main activity/main source of income is roughly equivalent to a full-time job. Accordingly, the 1968 sample is dichotomized into full-time paid employment and all other employment statuses, and in order to make comparisons, this division is also applied to the 1975–1991 surveys.

Long-standing diseases as an indicator of ill-health

A necessity when studying health trends is to find comparable indicators of health. In the five surveys, only one aspect of health is measured in a way that satisfies this requirement—the prevalence of long-standing diseases. The surveys asked about the respondents' present medical conditions. Up to 9 different diseases were recorded per respondent, and they were coded according to the International Classification of Diseases [56] and divided into long-standing and other diseases. From this information, the variable 'number of long-standing diseases' has been formed.

The wording of the questions were not completely identical in the five surveys (*cf* the documentation in the reports [50–54]), and the categorization of diseases as long-standing differed somewhat—it followed the classification scheme of the Central Bureau of Statistics in 1968, 1975 and 1985, but was based on the respondents' self-reports in 1987 and 1991. Nevertheless, the questions and procedures were sufficiently similar to believe that the variable measures the same type of health problems. The distributions of the variable supports this conclusion. In each of the surveys, around 50% report at least one long-standing disease, and the mean and standard deviation of the variable is fairly constant (Table 2). This stability can be taken as a confirmation of the reliability of the variable-assuming, of course, that no dramatic changes as regards the overall health situation among Norwegian women have occurred during the 70s and 80s.

It should be noted, however, that comparability is not only dependent upon the similarities between the survey instruments, but also on the social context. Changes in lay conceptions of health or more widespread contacts with the health services may lead to different answers, even if the underlying 'objective'

Table 2. Percent who have long-standing disease (LD). Marital, parental and employment statuses. Women 31–60 years, standardized by age, 1968–91

		1968	1975	1985	1987/1991
Marital status	Married	45.9	51.0	52.8	50.2
	Never married	45.9	57.3	53.7	48.8
	Divorced	65.9**	61.2*	65.1**	67.6**
	Widow	62.8**	66.1**	57.7	54.1
Children 0–16 years	None	50.9**	51.4	49.9*	50.0**
(only women aged 31–50)	One	43.4	47.6	52.9**	49.5**
	Two	40.0	45.6	50.1*	44.1*
	Three or more	37.6	48.3	40.7	34.1
Labour force participation	Full-time job	49.0	49.4	47.5	46.7
	Not full-time job	46.9	53.4	58.1**	55.7
Among those without full-time job (only 1975–1987/91):	Part-time		53.9	52.8*	51.6
	Homemaker		50.8	58.3**	51.0
	Others		71.2**	73.7**	71.6**
Total[a]		48.4	53.4	53.0	50.4
Average No. of LDs[a]		0.92	0.94	0.90	0.89
Standard deviation		1.28	1.17	1.16	1.27
(*N*)		(2082)	(1981)	(1902)	(1879)

* = $P < 0.05$ ** = $P < 0.01$. Significance refers to the gap between the actual category and the reference categories married, three or more children, and full-time employed.
[a]Not standardized by age.

health has remained unchanged. To estimate the extent of such changes during the study period is difficult, and this constitutes one uncertainty in this study.

Number of long-standing diseases is a measurement of morbidity which focusses on disease. It describes morbidity not by asking about personal experiences of one's health, but by reporting the names of medical conditions. It is probable that the information about diagnoses gathered in the surveys is, by and large, compatible with a medical view. Almost all of the long-standing diseases which are reported have also been brought to the attention of a physician, and usually, the respondents repeat the name of the disease they were told by their doctor. A study of the validity of the method conducted in connection with the 1968 survey revealed that the respondents' physicians could, in most cases, confirm the information given in the interviews [57].

Thus, number of long-standing diseases represent health as medically confirmed conditions, and in this sense it describes an important part of a person's morbidity. It is, however, one-sided. Acute conditions and short-term diseases, symptoms, vague mental problems, functional limitations and self-assessed health are not measured by the variable. The graveness of a person's health problems is certainly related to her number of long-standing diseases, but nevertheless it is a weakness that the variable counts each reported condition as one disease, irrespective of its seriousness. Moreover, the variable is not very sensitive since about half of the respondents report no long-standing disease, and among this 'healthy' half, no variation in health is measured.

The variable is correlated to some extent to other measurements of morbidity. In the 1985 survey, which measured ill-health in several ways, the correlation (Pearson's r) between number of long-standing diseases and self-assessed health status (five-point ordinal scale) was 0.43. The correlation with an index counting practical and social problems caused by ill-health was 0.38, and 0.32 with a mental health index based on the Hopkins Symptom Check List [58]. Accordingly, one may conclude that number of long-standing diseases is an important indicator of morbidity in a disease-oriented sense, and besides it is an approximate indicator of the respondents' experience of ill-health in a more all-round sense.

ANALYSIS

Changes in the relationship over time? Bivariate analyses

In each survey, about 50% of the samples have at least one long-standing diseases, and the average number of long-standing diseases are about 0.90 (Table 2). The variation between the surveys may be due to 'real' developments, such as actual changes in health or in respondents' reactions to questions about health. It is also possible that changes in the age composition of the population, differences in measurement methods, or sample bias, have influenced the results. The variation is small, however, and a likely inference is that the occurrence of long-standing diseases among women 31–60 years of age has remained more or less the same during this period.

The distribution of long-standing diseases is very skewed: about half of the women have no long-standing disease, and about 10% have three or more. This is a considerable deviation from a normal distribution, and there is a possibility that different summary measures will lead to different interpret-

ations. Therefore, the associations are described both by percentages having at least one long-standing disease (Table 2) and by average number of long-standing diseases (Fig. 1—here some statuses are combined, and some others are not included). As can be seen, the general pattern is quite similar between the two ways of data presentation.

The occurrence of long-standing diseases increases markedly by age. In the five samples taken together, 43.1% of respondents aged 31–40 have at least one, as against 61.4% among those aged 51–60. In Table 2 and Fig. 1, the effect of age has been eliminated by standardization. In each survey the respondents of each category has been weighted so that the age distribution on 10-years bands is identical to the age distribution of the total samples. When analyzing parental status, only women aged 31–50 are included, because there are very few women in their 50s who still have three, or even two, children aged 16 or younger. When omitting women 51–60 years of age from this crosstabulation, the confounding effects of mixing together, in the childless category, women who never have had any children with middle-aged mothers whose children are older than 16 are to some extent avoided.

As the samples of the two last surveys, 1987 and 1991, are only half as large as the previous surveys, they have been pooled in these analyses in order to increase the stability of the estimates.

The general pattern of Table 2 and Fig. 1 is not unexpected: previously married women have long-standing diseases more often than married women; childless women more often than mothers; and non-employed women more often than employed. It should be noted that the relatively detailed classification used in Table 2 provides information which is concealed when a more crude classification is used. When respondents are divided dichotomously into married and not married, the health differences within the not married category, between never married and previously married women, are overlooked. As can be seen from Table 2, in terms of long-standing diseases the health of never married women is very similar to that of married women. In a similar way, Table 2 shows very marked differences within the category of non-employed women, between homemakers and 'others'.

When interpreting these data, a difficult question is to distinguish between sample fluctuations and real trends. We should look for changes which are both significant and consistent during the whole period. We expect that the general health situation does not change fast, at least if there are no obvious reasons such as war or famine. Therefore, large fluctuations 'up' and 'down' during this period are not credible.

The first hypothesis stated above suggested smaller differences linked to marital status, but Table 2 and Fig. 1 do not show any distinct changes. It could be tempting to interpret Table 2 as indicating that health disadvantages of widowed women have been reduced in the 1980s. However, there are few widows among women aged 31–60 (*cf* Table 1) and the estimates have large confidence intervals. It seems more appropriate to combine divorced and widowed women into a category of previously married women. Figure 1 indicates that their health disadvantages are approximately the same in the late 1980 as in the early 1970s—the smaller difference in the 1985 survey could be an effect of random sample variation.

The second hypothesis—stable associations between health and parental status—is not contradicted by Table 2 and Fig. 1. Health in terms of long-standing diseases are associated to the number of children (exception: 1975), but no distinct trend as regards these differences is visible.

As to the third hypothesis, the bivariate analyses support it. When women are divided into full-time employed and all other employment statuses, there appears a steady widening of the disparities—they are small in the 1968 survey but increase to a marked and significant inequality in the surveys from the late 1980s. The same picture is shown in Fig. 1 as to differences between employed and non-employed (this division is only possible from the 1975 survey and onwards), which were relatively small in 1975, but larger in the 1987/91 data.

It should be noticed, however, that Table 2 shows that homemakers are not much disadvantaged in comparison with employed women, while the category of 'others' are much more afflicted by health problems. There are no distinct changes 1975–1987/91 as regards these differentials. Nevertheless, the overall result is that disparities between employed and non-employed women have increased markedly in this period (Fig. 1). One reason for this development is that the proportion of homemakers has decreased while the proportion of 'others' has increased during these years (Table 1).

Multivariate analyses

The analyses of Table 2 and Fig. 1 focus on one particular status at a time, but we would be more confident in the results if multivariate analyses controlling for other variables display the same pattern. The OLS multiple regression analyses displayed in Table 3 provide material for a more rigorous treatment of the three hypotheses. In order to neutralize some sample fluctuations, the samples are combined into two groups: 1968–1975 and 1985–87–91. Thus, the analyses highlight possible contrasts between the early 70s and late 80s. The dependent variable is number of long-standing diseases. Marital status is represented by dummy variables, using married as reference category. Divorced and widowed are collapsed into previously married. The variable number of children is treated as an interval scale. Employment status is coded '0' when the respondent has a full-time job and '1' otherwise.

If the linkages between health and status have been altered since the early 70s, we would expect significant changes in the unstandardized regression coefficients

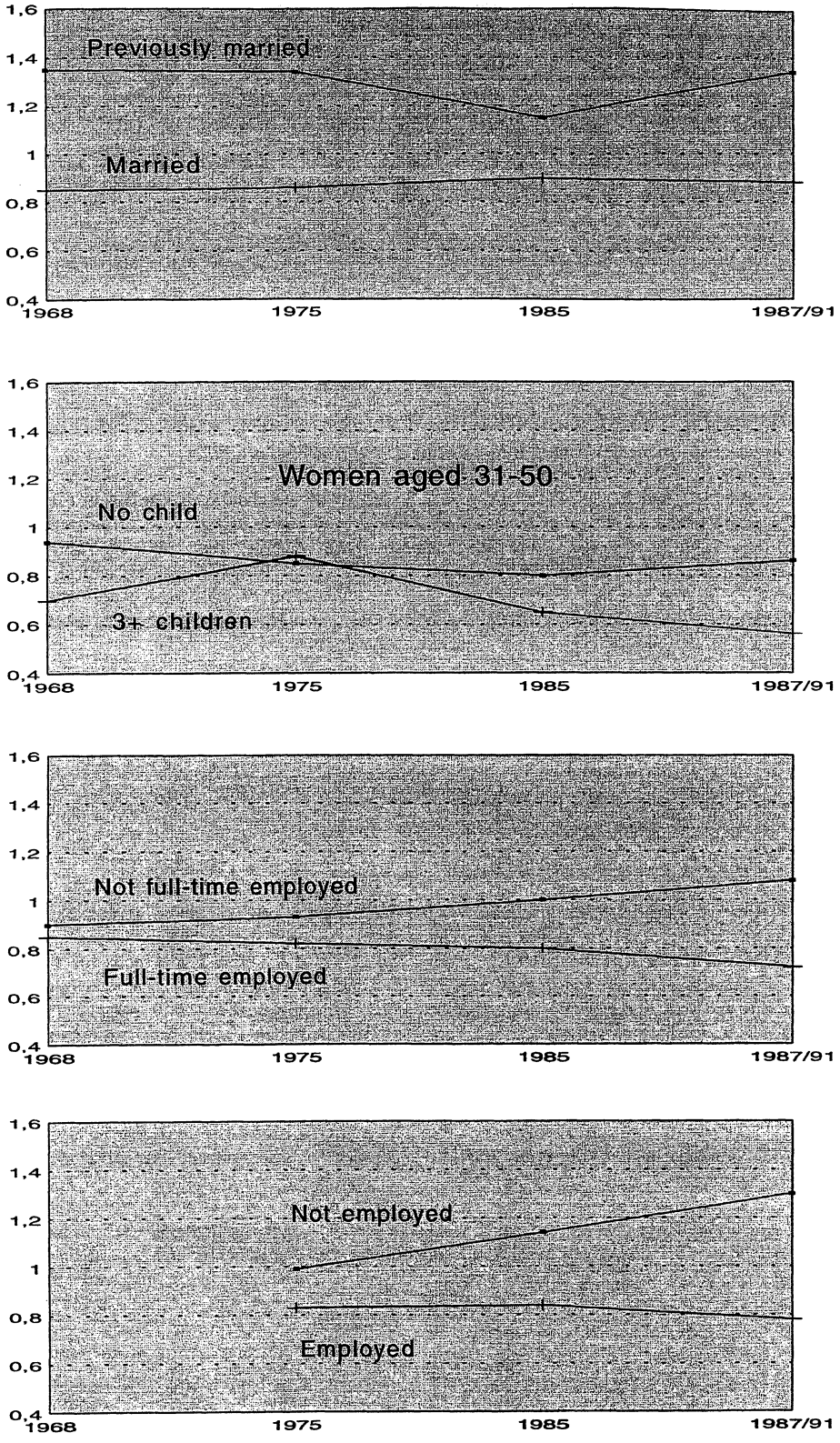


Fig. 1. Average number of long-standing diseases. Women aged 31–60, marital, parental and employment status, selected categories, 1968–1987/91. Age adjusted.

Table 3. Changes in the effects of age, marital, parental and employment status on number of long-standing diseases, between the early 70s and late 80s. Multiple regression analyses. Women aged 31–60

	Early 70s			Late 80s			Change		
	B_1	SE	*P*-val	B_2	SE	*P*-val	$B_2 - B_1$	SE	*P*-val
Age (year)	0.022	0.0029	0.000	0.016	0.0029	0.000	−0.006	0.0041	0.144
Never married	0.094	0.0741	0.204	0.081	0.0864	0.346	−0.013	0.1138	0.912
Previously married	0.470	0.0684	0.000	0.352	0.0633	0.000	−0.118	0.0932	0.102
Ref.cat.: married									
No. of children	−0.058	0.0229	0.011	−0.093	0.0247	0.000	−0.035	0.0337	0.298
Not full-time job	0.197	0.0446	0.000	0.319	0.0392	0.000	0.122	0.0594	0.020
Constant	−0.197			0.070					
Adj.R^2	0.052			0.058					
(*N*)	(4063)			(3781)					

B = unstandardized regression coefficient. SE = standard error of B. SE for change in B = the square root of $(SE^2_{\hat{B}_1} + SE^2_{\hat{B}_2})$. *P*-val = based on *t*-tests: probability of observed result if null hypotheses are true, one-sided for $B_2 - B_1$ for previously married and not full-time employment, otherwise two-sided.

(the Bs) between the early 70s and the late 80s data sets—in other words, we would expect the difference B_2–B_1 to be significantly different from zero (the columns to the right in Table 3). First, we note that after control for age, marital status and parental status, the effect of full-time employment is clearly higher in the late 80s than it was in the early 70s. Women without full-time employment had approx. 0.20 more long-standing diseases than full-time employed in the early 70s, and this difference had increased to 0.32 in the late 80s. The increase is significantly different from zero ($P = 0.020$, one-sided *t*-test), and the null hypothesis (H_0: no or reduced differences) can be rejected. Accordingly, the third hypothesis is confirmed by this analysis.

Neither is the second hypothesis, which predicted unchanged associations between number of long-standing diseases and parental status (number of children), contradicted by the results, as the difference $B_2 - B_1$ is not significant ($P = 0.298$). However, the null hypothesis is not 'proved' by the observation that differences do not reach significance level [59] (p. 16). An analysis of the statistical power of the test is necessary in order to evaluate the probability of making a Type II error, i.e. the error of not rejecting the null hypothesis, although the alternative H_1 which states that changes have occurred is in fact true. In order to do this analysis, one must decide a specific value for H_1. Thus, we have to answer the question: How large must the change be in the association between parental status and number of long-standing diseases to make us conclude that the 'no-change-hypothesis' is false? It is improbable that the association has remained *exactly* the same since the early 70s. But what constitutes a non-trivial change?

The answer to this question is to some extent arbitrary. I will consider two alternatives, which both constitute interesting deviations from the situation in the early 70s. First, I assume that the effect of parental status has been *doubled* between the early 70s and late 80s, in other words that $B_2 - B_1 = 2*B_1 - B_1 = -0.058$. At a 5% significance level, H_0 is not rejected if the observed difference $B_2 - B_1$ falls between ± 1.96 SEs of the sampling distribution of $B_2 - B_1$, that is, between −0.066 and +0.066. Using standard procedures (*cf* Ref.[60], pp. 269–286; [61], pp. 252–268), the calculation shows a probability of making a Type II error of 0.595, if $B_2 - B_1 = -0.058$. The statistical power of the significance test, given that the effect of parental status has in fact been doubled, is therefore approx. 40%.

The other alternative I consider, is that the association has changed to the extent that there is *no longer any effect of parental status* in the late 80s, in other words that $B_2 = 0$, and that H_1: $B_2 - B_1 = 0 - B_1 = +0.058$. Of course, in this case the calculation of statistical power duplicates the former calculation and again results in an estimate of approx. 40%.

From these calculations, given two specified values of $B_2 - B_1$ which both represent substantially changed effects of parental status, it appears that there is a considerable risk of making a Type II error. We cannot be sure that no change has occurred, as the test of statistical significance when confronted with two interesting alternatives has not more than about 40% statistical power, while we usually want tests which have a power of at least 70% [62].

Lastly, the first hypothesis, which expected diminished effects of marital status. There are very small differences between married and never married women, and therefore I will focus on the difference between currently married and previously married. The observed value $B_2 - B_1$ is −0.118. Thus, there is an observed change in the expected direction, but the change is not significant ($P = 0.102$, one-sided *t*-test). Accordingly, the hypothesis has weak support in the data. Also in this case, an analysis of statistical power is required in order to evaluate the null hypothesis that no change has occurred. The observed coefficient during the early 70s was as high as 0.47. Assuming that an interesting improvement in the health disadvantages of previously married has occurred if the difference has been reduced by *one third* between the early 70s and late 80s (thus, assuming the true situation to be H_1:$B_2 - B_1 = \frac{2}{3}*B_1 - B_1 = -0.157$), the statistical power of the test is about 52% (given a significance level of 5%). A conclusion that no reduction of the disadvantages of previously married has occurred is therefore weakly founded. But this argument depends of course on what reduction we consider important. If we decide that the disadvantages of previously

married women must have been *halved* between the early 70s and late 80s in order to say that the situation has been 'really' changed (H_1:$B_2 - B_1 = \frac{1}{2}*B_1 - B_1 = -0.235$), the power of the significance test is 81%, which is quite acceptable.

The multiple regression analyses of Table 3 control for age and women's statuses, but do not include interaction terms. In other words, the analyses are implicitly based upon the debatable assumption that the effect of each particular status is independent of women's other statuses. It is possible that combinations of statuses have health implications which are not predictable from the knowledge of the association between health and each single status—this has been the question behind the examination of the effects of multiple roles e.g. [9, 14, 16].

In order to test this possibility, several interaction terms have been included in the multiple regression analyses of Table 3. Because of multicollinearity, it was difficult to estimate the regression coefficients. Therefore, an alternative method was used: dummy variables representing each separate combination of statuses were analyzed by means of multiple regression. The results of these analyses, using married women with children and full-time employment as reference category, are displayed in Table 4.

The small increases in R^2 in the analyses of Table 4, compared to the previous analyses of Table 3, signify that the interaction effects are not large.

Interaction effects can also be examined by comparing the coefficients of specific combinations of statuses (Table 4) to the sums of the coefficients of each single status, displayed in Table 3.

According to the analyses of Table 3, the average difference between full-time employment and all other employment statuses was about 0.20 long-standing diseases in the early 70s. The difference between married and previously married was 0.47, and between childless women and mothers the difference was 0.20 (this is the coefficient for parental status, early 70s, using a no child/children dichotomy in the model analyzed in Table 3 instead of number of children).

Therefore, if there are no interaction effects in the early 70s, we would expect married childless women without full-time job to have about (0.20 + 0.20 =)0.40 more long-standing diseases than married full-time employed mothers. The actual difference when these combinations of statuses are analyzed in Table 4 is 0.33. In other words, this example does not show any sizeable interaction effect. On inspecting other coefficients closely in this way, we do not find much interaction effects. In particular, one should note that the analyses of Table 4 do not support the hypothesis about 'role overload'. In terms of long-standing diseases, the health of women who are married, who have children and who are full-time employed, is good compared to other combinations of statuses.

However, there are some striking interaction effects in the late 80s data set, which were not present in the early 70s. At that time, there were significant differences in health among full-time employed women, associated with their marital and parental status. For instance, both full-time employed married without children and full-time employed, previously married women (with or without children) had significantly more long-standing diseases than the reference category full-time employed, married mothers. These differences have practically disappeared in the analyses of the late 80s, when there are no significant inequalities among full-time employed women, no matter their marital and/or parental status. Accordingly, it seems that the effect of full-time employment has become much stronger. Earlier, full-time employment had no special impact on the effect of other statuses, but in the late 80s, its positive

Table 4. Effects of age and combinations of marital, parental and employment statuses (dummy variables) on number of long-standing diseases. Multiple regression analyses. Early 70s and late 80s. Women aged 31–60. Unstandardized regression coefficients. Reference category married, children and full-time employment

	Early 70s	(*N*)	Late 80s	(*N*)
Full-time job and:				
—married with children (ref.cat)		(381)		(694)
—married without children	0.185*	(383)	−0.089	(602)
—never married without children	0.127	(215)	−0.072	(111)
—previously married with children	0.304*	(71)	0.094	(75)
—previously married no children	0.590***	(108)	0.183	(112)
Not full-time job and:				
—married with children	0.157*	(1668)	0.077	(1174)
—married without children	0.330***	(948)	0.360***	(721)
—never married without children	0.659***	(81)	0.713***	(51)
—previously married with children	0.465**	(77)	0.698***	(66)
—previously married no children	1.074***	(98)	0.789***	(132)
Age	0.019***		0.017***	
Constant	− 0.201		0.016	
Adj.R^2	0.056		0.062	
(*N*)		(4030)		(3738)

*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$. Never married with full-time job/children, and never married women with children but without full-time job are excluded because of <30 respondents.

association with health overrides the potential negative effects of being previously married or childless.

CONCLUDING DISCUSSION

Main results

Three hypotheses were put forward in the earlier sections of this article. The first one concerned the associations between health and marital status, and stated that the health disadvantages of previously married women, compared to currently married, have become smaller during the last decades. In the multivariate analyses, the observed results were in the expected direction, but the *P*-value was only 0.102, and the null hypothesis of no change could not be rejected. However, if one assumes that the true situation is that there has been a reduction of these differences by one third, which can be considered a substantial change, the power of the significance test is only about 52%. This means that one cannot infer that no change has occurred with any high degree of certainty. A prudent conclusion as to this hypothesis is to suspend judgment until more evidence is available.

The data show that there is an inverse relation between number of children and health problems, and the second hypothesis suggested stability as to this association. This hypothesis is not rejected, as the analysis does not detect any statistical significant change in the effect of parental status on number of long-standing diseases. But again one cannot rule out the possibility that non-trivial changes have occurred, as the evaluation of this result shows that the possibility of Type II error may be substantial.

As regards the third hypothesis, which predicted that the disparities between employed and not employed women have increased during the 70s and 80s, the results support it. The null hypothesis of no change has only a probability of 0.020, which means that it can be rejected, assuming, of course, that there is no measurement bias in the data. Accordingly, it seems that there appears a widening health gap associated with employment status: among women in paid employment, especially full-time employed, there has occurred a relative improvement in health, compared to non-employed women. But non-employed women include two different categories: housewives and 'others'. The latter are usually recipients of welfare state benefits, and the widening differentials between employed and non-employed have probably to some degree developed because the category of 'others' who have especially many health problems, has grown, while full-time homemakers, whose health usually is only a little worse than the health of employed women, make up a decreasing proportion among women.

The analysis of multiple roles reveals only small interaction effects of combinations of specific marital, parental and employment statuses. An exception is the emerging tendency of a protective effect of paid employment. Although there usually is a negative health effect associated with being previously married and/or childless, data from the late 80s show that these negative effects disappear when women who have these 'negative' statuses are also full-time employed. This interaction effect was not visible in the early 70s, and this may be an indication that paid employment has acquired a more important part in women's lives: when they 'succeed' as regards this aspect, marital break-ups or having no children are not associated with deteriorated health in the same way as some twenty years earlier. This result is consistent with Arber's observation, based on U.K. data from the middle of the 80s, that "paid employment is 'protective' of health, even if women have other characteristics which usually confer health disadvantage" [16] (p. 434).

Discussion

Thus, the hypothesis that inequalities related to marital status have diminished found weak support in the data, while the hypothesis about increasing differences associated with employment status is in accord with the analyses. In this concluding discussion I will focus on these results. How can they be interpreted? The data themselves do not include variables which allow us to study directly the causal mechanisms which are involved, and how social change may have modified their operations. We have to use statistical and historical knowledge from other sources in an attempt to construct a plausible 'story' which agrees with the results. Several relevant points have been made above, in the sections which preceded the data analysis. Here I will supplement that discussion.

The analyses of this article suggest that the causal mechanisms which generate inequalities in health related to marital status operate to a considerable extent independently of macro-social change. Both factors connected to the economic and ideological order of society, such as material disadvantages and low social esteem, as well as factors which are more located in the private sphere, such as depression, hopelessness, and feelings of loss and personal failure, have been proposed as potential causes for the health disadvantages of previously married women. In addition, it has been argued that health problems may increase the risk of being divorced—in other words, that health selection is involved. There have been rather dramatic changes in the institution of divorce during the last decades—changes which probably have improved to some extent the situation of divorced women both materially and in terms of social esteem. In spite of this, the health inequalities between married and divorced women have probably not been very much reduced—data suggest some reduction, but the change does not reach statistical significance. A reasonable inference is therefore that important causal

reasonable inference is therefore that important causal mechanisms are located in the private sphere, out of reach, so to speak, of macro-social change. An important factor may be psychic distress connected to the marital break-up itself, and there is also a possibility that health selection is involved, and neither of these causal mechanisms are easily influenced by social change at the macro level.

The increasing inequalities in health related to employment status, however, must probably be interpreted in terms of how social change has interacted with various causal mechanisms. In the late 80s, full-time working married women usually have fewer children and better daycare facilities, they do less housework, and they have shorter working hours and somewhat improved working conditions. Accordingly, the potential for negative health effects of this combination of statuses has probably decreased, and this is one reason why the difference between employed and non-employed women has widened. Ideological developments point in the same direction: to be 'only' a homemaker is not bringing much prestige in the late 80s, while being full-time employed is much more in line with prevailing cultural ideals, and this may be a contributing cause to the increased protective effects of employment suggested by these data.

In addition to these developments, which concern the operation of social causation mechanisms, one should also consider the part played by health selection. It is possible that good health has become a more explicit criterion when women enter paid employment. Furthermore, health related selection out of the labour market is certainly important. The development of various welfare state benefits has constructed a state regulated selection mechanism, which locates women whose health limits their working capacity into a special category of non-employed women. This mechanism, often termed 'the healthy worker effect', has become more important as a result of the growth of these benefits. Thus, a byproduct of the welfare state is a tendency of larger health differences between employed and non-employed women, because of health selection. (But it should of course not be forgotten that health related selection out of the labour force often has been preceded by social causation processes: low-status working women, who often experience poor working conditions, are over-represented among exits from the labour market.)

The question about inequalities associated with employment status has to some extent acquired a new meaning in countries such as Norway. This question came into focus when paid employment among women grew dramatically during the 60s and 70s. At that time, the relevant comparison was between employed women and housewives. Interest in this comparison was spurred by the heated debate about women's proper place, and proponents of women's participation in paid work were eager to point out that employed women had better health than housewives. In countries such as Norway, social change has to some extent made this debate obsolete. Around 1990, only about one out of ten women was a full-time homemaker (Table 1), and to be a housewife during one's entire adult life is quite unusual. Accordingly, the question about which activity is most beneficial for women's health—paid employment or housework?—is becoming more difficult to study because lifelong, full-time homemakers are rare. Women classified as homemakers in a cross-sectional survey are to a large extent women who had paid employment some time ago, and who will later re-enter employment.

When the health of employed and non-employed women is compared, the latter category was dominated by housewives in the early 70s, but by recipients of welfare state benefits in the late 80s. The comparison in the late 80s is in this sense not entirely comparable to the comparison made in the early 70s. This should be observed when discussing the findings that inequalities between employed and non-employed women have increased over time. Usually, increased health inequalities related to social positions are condemned, but in this case the increased inequalities, at least in part, are linked to the development of institutions which give women who suffer from health problems an opportunity to some income at the same time as they are relieved from the burdens of paid employment. One may object to this system of channeling people with health problems out of 'meaningful' activity, but it is clear that increased inequalities generated in this way cannot be judged in the same manner as we usually judge inequalities in health related to social positions.

Lastly, some general remarks about inequalities in health associated with social positions. Many have had the hope that developments during the second half of this century—such as the overall improvement in standard of living, the elimination of many unhealthy working milieus, the general spread of health information, and the development of health services—should gradually eliminate many socially patterned health inequalities. This article, together with many other studies in the field, indicate that this hope is not well founded. Inequalities associated with women's statuses tend to persist, and in some ways they even increase. The preconditions for breaking this tendency could be an interesting topic for future studies.

Acknowledgements—The survey data used in this paper have been delivered by the Norwegian Social Science Data Services. Data were collected and processed by the Central Bureau of Statistics (Norway). Neither of these institutions is responsible for the data analysis or for the interpretations provided in this article. The author thanks Espen Dahl, Dag Album, Natalie Rogoff Ramsøy, and three anonymous reviewers for useful comments.

REFERENCES

1. Gove W. R. Gender differences in mental and physical illness: the effects of fixed roles and nurturant roles. *Soc. Sci. Med.* **19**, 77, 1984.

2. Gove W. R. and Geerken M. The effects of children and employment on the mental health of married men and women. *Social Forces* **56,** 67, 1977.
3. Hibbard J. H. and Pope C. R. Employment characteristics and health status among men and women. *Women Hlth* **12,** 85, 1987.
4. Hibbard J. H. and Pope C. R. Effect of domestic and occupational roles on morbidity and mortality. *Soc. Sci. Med.* **32,** 805, 1991.
5. Hibbard J. H. and Pope C. R. Women's employment, social support, and mortality. *Women Hlth* **18,** 119, 1992.
6. Muller C. Health and health care of employed women and homemakers: family factors. *Women Hlth* **11,** 7, 1986.
7. Nathanson C. Social roles and health status among women: the significance of employment. *Soc. Sci. Med.* **14a,** 463, 1980.
8. Sorensen G. and Verbrugge L. M. Women, work, and health. *A. Rev. Publ. Hlth* **8,** 235, 1987.
9. Verbrugge L. M. Multiple roles and physical health of women and men. *J. Hlth Soc. Behav.* **24,** 16, 1983.
10. Verbrugge L. M. Role burdens and physical health of women and men. *Women Hlth* **11,** 47, 1986.
11. Verbrugge L. M. and Madans J. H. Social roles and health trends of American women. *Milbank Memorial Fund Q./Hlth Society* **63,** 691, 1985.
12. Waldron I. Employment and women's health: an analysis of causal relationships. *Int. J. Hlth Serv.* **10,** 435, 1980.
13. Waldron I. and Herold J. Employment, attitudes toward employment, and women's health. *Women Hlth* **11,** 79, 1986.
14. Waldron I. and Jacobs J. A. Effects of multiple roles on women's health—evidence from a national longitudinal study. *Women Hlth* **15,** 3, 1989.
15. Arber S. Opening the 'Black Box': Inequalities in Women's Health. In *New Directions in the Sociology of Health* (Edited by Abbott P. and Payne G.). The Falmer Press, London, 1990.
16. Arber S. Class, paid employment and family roles: making sense of structural disadvantage, gender and health status. *Soc. Sci. Med.* **32,** 425, 1991.
17. Arber S., Gilbert G. N. and Dale A. Paid employment and women's health: a benefit or a source of role strain? *Sociol. Hlth Illness* **7,** 375, 1985.
18. Arber S. and Lahelma E. Inequalities in women's and men's ill-health: Britain and Finland compared. *Soc. Sci. Med.* **37,** 1055, 1993.
19. Bartley M., Popay J. and Plewis I. Domestic conditions, paid employment and women's experience of ill-health. *Sociol. Hlth Illness* **14,** 313, 1992.
20. Hunt K. and Annandale E. Just the job? Is the relationship between health and domestic and paid work gender-specific? *Sociol. Hlth Illness* **15,** 632, 1993.
21. Popay J., Bartley M. and Owen C. Gender inequalities in health: social position, affective disorders and minor physical morbidity. *Soc. Sci. Med.* **36,** 21, 1993.
22. Næss S. Yrkeskvinne—husmor? Gifte kvinners livskvalitet. (Employed woman—housewife? The quality of life of married women.) *INAS-Report* 86, **3,** Oslo, 1986.
23. Søgaard A. J. Er det egentlig sunt å jobbe? (Is work really healthy?) In *Kvinnemedisin* (*Women's Medicine*) (Edited by Schei B., Botten G. and Sundby J.), p. 221. Ad Notam–Gyldendal, Oslo, 1993.
24. Adelmann P. K., Antonucci T. C., Crohan S. E. and Coleman L. M. A causal analysis of employment and health in midlife women. *Women Hlth* **16,** 5, 1990.
25. Arber S. and Lahelma E. Women, paid employment and ill-health in Britain and Finland. *Acta Sociol.* **36,** 121, 1993.
26. Houston B. K., Cates D. S. and Kelly K. E. Job stress, psychosocial strain, and physical health problems in women employed full-time outside the home and homemakers. *Women Hlth* **19,** 1, 1992.
27. Jennings S., Mazaik C. and McKinlay S. Women and work: an investigation of the association between health and employment status in middle-aged women. *Soc. Sci. Med.* **19,** 423, 1984.
28. Kandel D. B., Davies M. and Raveis V. H. The stressfulness of daily social roles for women: marital, occupational and household roles. *J. Hlth Soc. Behav.* **26,** 64, 1985.
29. Lennon M. C. and Rosenfield S. Women and mental health: the interaction of job and family conditions. *J. Hlth Soc. Behav.* **33,** 316, 1992.
30. Rosenfield S. The effects of women's employment, personal control and sex differences in mental health. *J. Hlth Soc. Behav.* **30,** 77, 1989.
31. Shehan C. L. Wives' work and psychological well-being: an extension of Gove's social role theory of depression. *Sex Roles* **11,** 881, 1984.
32. Blane D., Davey Smith G. and Bartley M. Social selection: what does it contribute to social class differences in health? *Sociol. Hlth Illness* **15,** 1, 1993.
33. Bloor M., Samphier M. and Prior L. Artefact explanations of inequalities in health: an assessment of the evidence. *Sociol. Hlth Illness* **9,** 231, 1987.
34. Wilkinson R. G. Socio-economic differences in mortality: interpreting the data on their size and trends. In *Class and Health. Research and Longitudinal Data* (Edited by Wilkinson R. G.), p. 4. Tavistock Publications, London, 1986.
35. West P. Rethinking the health selection explanation for health inequalities. *Soc. Sci. Med.* **32,** 373, 1991.
36. Lundberg O. Childhood living conditions, health status, and social mobility: a contribution to the health selection debate. *Eur. Sociol. Rev.* **7,** 149, 1991.
37. Schwalbe M. L. and Staples C. L. Class position, work experience, and health. *Int. J. Hlth Serv.* **16,** 583, 1986.
38. La Rosa J. H. Women, work, and health: employment as a risk factor for coronary heart disease. *Am. J. Obstetr. Gynecol.* **158,** 1597, 1988.
39. Dahl E. Social inequality in health—the role of the healthy worker effect. *Soc. Sci. Med.* **36,** 1077, 1993.
40. Dahl E. and Kjærsgaard P. Social mobility and inequality in mortality. An assessment of the health selection hypothesis. *Eur. J. publ. Hlth* **3,** 124, 1993.
41. Mastekaasa A. Psychological well-being and marital dissolution: selection effects? *J. Family Issues* **15,** 208, 1994.
42. Mastekaasa A. Ekteskap og psykisk helse-mulige årsaksmekanismer. (Marriage and mental health—possible causal mechanisms.) *Working Paper* **91: 6,** Institute of social research, Oslo, 1991.
43. Central Bureau of Statistics, Norway. *Sosialt utsyn* 1993 (Social trends 1993.) Konsvinger, Oslo, 1993.
44. National Insurance Administration, Norway. *Årsmelding og regnskap for 1971* (Annual Report 1971.). Oslo, 1971.
45. National Insurance Administration, Norway. *Trygdestatistisk årbok 1993* (Social Insurance Statistical Yearbook 1993.), Oslo, 1993.
46. Domar A. D., Zuttermeister P. C. and Friedman R. The psychological effect of infertility—a comparison with patients with other medical conditions. *J. Psychosomatic Obstetr. Gynaecol.* **14,** 45, 1993.
47. Central Bureau of Statistics, Norway. Statistical Yearbook 1992. *NOS C 32.* Kongsvinger, Oslo 1992.
48. Tønnesen C. U. Unpublished manuscript. Institute of Applied Social Research, Oslo, 1994.
49. Haraldsen G. and Kitterød H. *Døgnet rundt. Tidsbruk og tidsorganisering 1970–90.* (Around the clock. Time-use and time-organization 1970–90.) Central Bureau of Statistics, Norway, Oslo-Kongsvinger, 1992.

50. Central Bureau of Statistics, Norway. Health Survey 1968. *NOS A 321*, Oslo, 1970.
51. Central Bureau of Statistics, Norway. Health Survey 1975. *NOS A 894*, Oslo, 1977.
52. Central Bureau of Statistics, Norway. Health Survey 1985. *NOS B 692*. Kongsvinger, Oslo, 1987.
53. Central Bureau of Statistics, Norway. *Survey of Level of Living 1987. NOS B 772*. Kongsvinger, Oslo, 1988.
54. Central Bureau of Statistics, Norway. *Survey of Level of Living 1991. NOS C 43*. Kongsvinger, Oslo, 1992.
55. Rogoff Ramsøy N. Non-marital cohabitation and change in norms. *Acta Sociol.* **37**, 23, 1994.
56. Central Bureau of Statistics, Norway. Standard for gruppering av sykdommer-skader-dødsårsaker. *SSH 24*, Oslo, 1973.
57. Bjerkedal T. and Bakketeig L. Påliteligheten av intervjuopplysninger om helseforhold. (Reliability of health information obtained in interviews.) *Tidsskrift Den norske lægeforening* **95**, 927, 1975.
58. Derogatis L. R., Limpman R. S., Rickels K., Uhlenhuth E. H. and Covi L. The Hopkins Symptom Checklist (HSCL). A measure of primary symptom dimensions. In *Psychological Measurements in Psychopharmacology* (Edited by Pichot P.). Karger, Basel, 1974.
59. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. Academic Press, New York, 1977.
60. Hays W. L. *Statistics for Psychologists*. Holt, Rinehart and Winston, New York, 1963.
61. Wonnacott T. H. and Wonnacott R. J. *Introductory Statistics*, 3rd edn. John Wiley and Sons, New York, 1977.
62. Denenberg V. H. Foreword. In *How Many Subjects? Statistical Power Analysis in Research* (Edited by Kraemer H. C. and Thiemann S.). Sage Publications, Newbury Park, 1987.

Paper IV:

"Employment status and women's health – exploring the dynamics", *Acta Sociologica*, 38 (3): 231-249, 1995

Acta Sociologica (1995) 38:231–249

Employment Status and Women's Health – Exploring the Dynamics

Jon Ivar Elstad
Institute of Applied Social Research, Oslo

Employed women report fewer health problems than non-employed women. The social causation explanation argues that these inequalities arise because employment statuses differ as to their consequences for health, while the health-related mobility explanation claims that differences emerge because of selective mobility. The present study focuses on the dynamic relationships between health and employment statuses. Panel data from Norwegian surveys 1980–91 are analysed. Differences in health are already present when young women enter employment statuses. This pattern is reproduced through numerous employment status transitions and changes in health. Log-linear analyses indicate that health-related mobility is more important than social causation, although both models are relevant. Further analyses indicate that health-related mobility primarily occurs in terms of health problems that cause reductions in paid employment.

Correspondence to: Jon Ivar Elstad, Institute of Applied Social Research, Munthesgt. 31, N-0260 Oslo, Norway.

1. Introduction

The growth in female employment was accompanied by warnings that women's 'double work', as both housewives and paid employees, would endanger their health. The bulk of empirical research indicates, however, that the health of employed women is favourable compared with that of women who do not have paid work. This was immediately evident when research started on the topic, and almost twenty years ago, a prominent researcher within the field, C. Nathanson, wrote that 'employment has perhaps the most clearly positive effects on women's health of any variable investigated to date' (quoted by Bartley et al. 1992:315). The results of later research are generally in accord with these early findings. Studies in the US (for example, Gove & Geerken 1977; Nathanson 1980; Waldron 1980; Verbrugge 1983, 1986; Hibbard & Pope 1987), in the UK (for example, Arber et al. 1985; Arber 1990, 1991; Bartley et al. 1992), and in the Nordic countries (Næss 1986; Axelsson 1992; Sogaard 1993; Arber & Lahelma 1993a, 1993b; see also Table 1) provide further confirmation of the favourable health of employed women, as compared with that of their non-working 'sisters'. Most studies are restricted to self-reported morbidity, but there are also studies which find that employment status predicts women's mortality (Passannante & Nathanson 1985; Hibbard & Pope 1991; Smith & Waitzman 1994; Vågerö 1994).

How is this pattern generated? Why do these inequalities in women's health emerge between those who are employed and those who are not? In most studies two contrasting views are usually discussed: (1) the social causation model, which states that these differences emerge because it is more healthful to have paid employment than to be a homemaker, and (2) the health-related mobility model, which claims that these differences are produced because of selective mobility processes which tend to locate women with better health in paid work.

The article first gives an overview and some comments on these two explanatory models. Thereafter, Norwegian survey data are used in order to investigate these questions. Both models mentioned above are dynamic in the sense that they assume that

Table 1. *Limiting long-standing illness and psychological problems (per cent). Women aged 16–67, four employment statuses. Surveys of Level of Living, Norway, 1980–83–87–91, separately and pooled.*

		Employed				
		Full-time	Part-time	Home-maker	Pension	Total
1980						
	Limiting long-st. illness	20.7	26.3*	30.9***	72.8***	29.6
	Psychological problems	32.2	35.6	38.4*	68.4***	38.0
	(N)	(552)	(472)	(450)	(136)	(1610)
1983						
	Limiting long-st. illness	23.9	22.3	36.8***	60.2***	31.0
	Psychological problems	30.3	36.3*	39.2**	61.2***	38.1
	(N)	(544)	(471)	(337)	(206)	(1558)
1987						
	Limiting long-st. illness	17.5	26.4***	25.0*	63.6***	26.6
	Psychological problems	30.1	37.9**	40.4**	64.2***	37.8
	(N)	(724)	(470)	(208)	(187)	(1589)
1991						
	Limiting long-st. illness	17.6	23.8*	35.7***	60.3***	29.7
	Psychological problems	29.5	30.7	35.7	52.5***	35.2
	(N)	(664)	(361)	(140)	(305)	(1470)
1980–91 pooled						
	Limiting long-st. illness	19.6	24.7***	32.2***	63.1***	29.2
	Psychological problems	30.5	35.4***	38.7***	59.8***	37.3
	(N)	(2484)	(1774)	(1135)	(834)	(6227)

Note: Students not included. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Statistical significance refers to the category compared to full-time employed.

the distribution of health problems is the result of processes which have evolved over time. Many studies which investigate these differences are based on cross-sectional data, which do not allow for a more thorough analysis of the dynamic processes involved. The survey material analysed here also includes panel data which are more appropriate for examining these questions. The aims of the analyses are to describe inequalities in women's health associated with their employment statuses in Norway during the 1980s, and to explore the processes which generate these inequalities. In particular, an attempt is made to evaluate the relative contribution of social causation and health-related mobility, respectively, in producing the observed differences.

2. Explanatory models

2.1. The social causation model: does employment benefit women's health?

The essence of the social causation model in explaining inequalities in health is that social positions differ with respect to their material and psychosocial circumstances. In consequence of this, the resulting health of the incumbents of these social positions also tends to vary (Townsend & Davidson 1982; Arber et al. 1985; Lundberg 1990, 1991; Dahl 1994a). As regards inequalities in women's health, this model suggests a research strategy which tries to discover the crucial factors that vary between the typical situations of employed versus non-employed, and that are capable of generating health differentials.

It has been hypothesized that paid employment could endanger women's health. This possibility has been investigated, for instance by examining the hazards connected with unhealthy working milieux (Waldron 1980; Sorensen & Verbrugge 1987), or by exploring possible adverse effects on working women's health from overwork and frustration connected with the combination of multiple roles as mother, spouse and employee (Verbrugge 1983; Arber et al. 1985; Bartley et al. 1992). But since the majority of studies suggest that women who have paid work are not suffering from worse, but rather enjoying better health than non-working women, it seems that any possible detrimental effects of paid employment are generally overshadowed by other factors.

Accordingly, much research inspired by this model has been directed at discovering factors encountered by working women which enhance positive health, and which distinguish their situation from circumstances commonly experienced by non-working women (usually housewives). One hypothesis is called the 'role enhancement model' (Bartley et al. 1992), which supposes that health benefits from participation in socially valued activities, and therefore the combination of marriage, parenthood and employment – all socially valued roles – tends to correlate with good health. More specific hypotheses examine how the situation of employed women, compared with that of homemakers, entails factors which could be advantageous to health: personal income and the feeling of power and independence which follows from it (Rosenfield 1989), opportunities for self-actualization, and friendship and social support from work-mates (Nathanson 1980). Homemakers, on the other hand, can be burdened by monotonous housework and child care, with few challenges and little opportunity for self-actualization, and they may suffer from few social contacts with adults (Gove & Geerken 1977; Gove 1984). Thus, it is hypothesized that paid work is generally a source of gratification, which engenders self-reliance, self-confidence, and other positive feelings which, according to the psychosomatic disease model, are beneficial to good health and help people avoid and overcome health problems (Gerhardt 1989:264f.).

2.2. Health-related mobility: selection for health

The model which explains inequalities in women's health as a result of health-related mobility is in many respects the opposite of the social causation model. The social causation model focuses on the health consequences of experiencing different social positions, while the health-related mobility model directs attention to recruitment patterns into these social positions (Illsley 1986; Lundberg 1991; West 1991; Blane et al. 1993; Dahl 1994b) – thus, women in some employment statuses have more health problems because their health was relatively unfavourable at the time they entered these employment statuses.

Two types of health-related mobility are commonly discussed. The first one is often termed (direct) health selection, implying that healthy individuals are somehow 'selected' into particular social positions – often the most rewarding ones – while sick people are pushed into, or allotted to, other, usually less fortunate, positions. Especially as regards social class inequalities in health, health selection has been associated with social Darwinism (West 1991; Dahl 1994b), because the underlying idea is that the most fit in terms of health are appointed to the most demanding, but also the most powerful and gratifying, social positions. Perhaps to avoid this affinity to social Darwinist ideas and their biological and individualistic bias, some researchers prefer the term 'social selection' (Blane et al. 1993). This term emphasizes that selection is a social process – various actors and agents are involved in a social interplay influenced by structures, norms, power, and other societal factors (West 1991).

With respect to associations between health and women's employment statuses, it is obvious that health (or social) selection generates health differentials when women become disability pensioners, or start receiving other kinds of public benefits granted on health criteria (Dahl 1994b). This process may be termed 'state-regulated health selection'. Because of the increase in recipients of welfare state ben-

efits during recent decades (Statistics Norway 1995), it is likely that health selection in this sense has increased during the 1970s and 1980s.

Another type of health selection is employer discrimination. When employers scrutinize applicants, they may select women who are in good health. Among women already employed, those who have health difficulties may be pressurized out of work. Women's self-constraint can have the same effect: on deciding whether or not to seek employment, to have full-time or part-time work, or to continue work if already employed, a woman may consider her health situation and refrain from employment if her health makes it inconvenient or strainful. But self-constraint in this sense can often interact with other factors. Because of prior rejections by employers, a woman may decide to avoid such humiliating experiences again. If she quits paid employment, ostensibly because she thinks her health is suffering, the underlying cause may be that the employer has implemented harsh production norms which exclude all but the most healthy.

2.3. Health-related mobility: indirect selection

The processes described above imply that health attributes are involved in women's movements into and out of employment statuses in a direct and manifest way. This need not be the case, however, and it has been pointed out that health inequalities can arise in consequence of mobility without any manifest or conscious selection of individuals because of their healthiness or sickness. A common term for this phenomenon is 'indirect health selection' (Wilkinson 1986:4; West 1991; Lundberg 1991; Blane et al. 1993; Dahl 1994b). The difference between direct and indirect selection is that, with regard to the former process, health attributes are causally related to subsequent movements between social positions, while in the latter process these movements are primarily affected by other causal factors. Mobility not caused by health may nevertheless generate inequalities in health, if the causal factors which effectuate mobility are also correlated with health. The association between health and mobility is in this case not a causal chain, but a spurious association arising because a common factor affects both (Waldron 1980; Lundberg 1991).

The association between health, employment status and education may be an example of an indirect selection process. The propensity to enter paid employment, in particular full-time employment, is connected to women's educational level (for Norwegian examples, cf. Central Bureau of Statistics 1993:131). Higher education may furthermore cause better health, if it leads to a healthier lifestyle or a more appropriate use of health services. Thus, the correspondence between (full-time) paid employment and better health is generated by a common background factor: education.

Other individual attributes may function in the same way. A high socio-economic background can, on the one hand, lead to better health, because of better nutrition in childhood (Forsdahl 1977; Barker 1991), and can on the other hand increase the probability of having full-time employment, either through the mechanism of higher education, or because employers prefer employees with a middle-class background. Thus, the correlation between full-time employment and health can develop through long-term causal chains which originate in early childhood. Furthermore, height may be connected to physical strength and healthiness and may at the same time be an attribute which enhances women's chances in the labour market (West 1991). To be a single mother is also a probable candidate as a common factor influencing both employment status and health: if single mothers tend to stay out of work because of their caring tasks, and their situation at the same time tends to lead to depression or other forms of health problems, health and employment status will correlate. The association between health problems and receiving a pension among women in their twenties (cf. Table 2B) could in part be such an indirect health selection effect, as some of these women are unmarried mothers or divorcees receiving single supporter benefits.

Table 2. *Limiting long-standing illness and psychological problems (per cent). A: Students aged 16–30 at first interview, by employment status at second interview. Two-wave panel data. B: Women aged 16–25. Pooled Surveys of Level of Living, Norway, 1980–91. C: Women aged 51–67. Pooled Surveys of Level of Living, Norway, 1980–91.*

	Employed		Home-maker	Pension	Total
	Full-time	Part-time			
A: Students at first interview					
Limiting long-st. illness	14.3	21.7	36.4*		19.7
Psychological problems	32.5	56.5*	50.0		40.2
(N)	(77)	(23)	(22)		(122)
B: Women aged 16–25					
Limiting long-st. illness	15.2	22.8*	21.3*	26.4*	18.4
Psychological problems	31.7	40.1*	41.0*	58.3***	36.6
(N)	(630)	(202)	(183)	(72)	(1087)
C: Women aged 51–67					
Limiting long-st. illness	27.3	31.0	43.3***	68.3***	44.0
Psychological problems	30.2	39.0**	42.2***	60.0***	43.9
(N)	(444)	(464)	(360)	(562)	(1830)

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Statistical significance refers to the category compared to full-time employed.

2.4. Supplementary remarks

Social causation and the two forms of health-related mobility – direct and indirect selection – are not the only models which have been proposed for explaining inequalities in health between social positions. It has been discussed whether reported inequalities are connected to methodological problems (the artefact explanation, cf. Townsend & Davidson 1982; Fox et al. 1986; Bloor et al. 1987), or whether inequalities develop as a result of complex interactional processes (Bartley 1991; Thebaud-Mony 1994). Nevertheless, social causation and health-related mobility are the dominating approaches, and they are the focus of this study.

Social causation is regarded as the main explanatory model for explaining differences in women's health-related employment statuses in many studies (cf., for instance, Gove & Geerken 1977; Nathanson 1980; Verbrugge 1983; Waldron & Jacobs 1989; Søgaard 1993), but it is also widely accepted that health-related mobility plays a part. A convincing argument for the relevance of the latter explanation is found in one of the few panel studies on this theme in the Nordic countries: Axelsson's investigation of Swedish women's transition to paid employment, which shows that the psychological health of housewives in 1968 was clearly related to their subsequent entrance or non-entrance into the labour market during the 1970s (1992:115–119).

As 'pure' models, social causation and health-related mobility are in opposition to each other, but in an actual society these types of processes can coexist. It is not difficult to find plausible examples that fit into each of the three main types of processes described above. Feelings of meaninglessness following early retirement may lead to psychological malaise (an example of social causation); women may choose homemaking instead of paid employment because of health problems (direct health selection); employers may prefer well-educated women, thereby unwittingly hiring healthy women (health-related mobility in the form of indirect selection), etc. Thus, a reasonable hypothesis is that the pattern of inequalities is generated through a com-

bination of social causation and health-related mobility.

In addition, one should note the possibility of historical change as regards the impact of these mechanisms. Social causation and health-related mobility are connected to the socio-political context (Arber & Lahelma 1993b; Elstad forthcoming). The health consequences of being in one rather than another employment status are, for example, related to the social prestige allotted to these statuses, which may change over time. Developments of rights of employees and the criteria used for granting welfare state benefits influence the scope of health selection. Therefore, findings in one society in a certain period about the impact of social causation and health-related mobility are not necessarily generalizable to other social contexts.

3. Modelling the relationships between health and employment status

3.1. Dynamics

The discussion above points to various dynamic processes which generate the pattern of health and illness associated with employment statuses. One question concerns the distribution of health among those who are recruited into 'the system'. In theory, the observed health differentials could simply be a consequence of differences present already at the time when young women entered their first employment status. Obviously, this initial distribution has some impact on the overall pattern of inequalities and should therefore be examined.

Moreover, the pattern of inequalities will be influenced by two types of transition: changes in health, and changes in employment status. A woman's health situation, whether good or bad, may be stable for many years; it may improve if 'old' health problems disappear; or it may deteriorate because new illnesses emerge. Corresponding types of transitions (or stability) will take place with respect to employment statuses: women may change employment status – either towards a higher level of employment participation (for instance, from part-time to full-time work), or to a lower level (for instance, from employment to homemaking).

Thus, inequalities at a specific point in time will be an aggregated result of the initial distribution and of the transitions and changes in health and employment status up to the observation point.

In general, social causation within these dynamics will consist of health developments which follow employment statuses – or employment status transitions – in time. This can take several forms: an unhealthy employment status may gradually cause new health problems; a healthy employment status may protect against the wear and tear of ageing; a change from an unhealthy to a healthy employment status may lead to improved health, and so on.

The mechanisms of health-related mobility will usually appear in terms of the opposite time-ordering – employment status will follow health in time. This may also occur in many specific ways: good health may have the consequence that a woman remains in a demanding employment status (for instance, full-time work); the existence of health problems may increase the probability that she later on reduces her working time; an improvement in health may be followed by a change of employment status, etc. But health-related mobility in the form of indirect health selection (cf. section 2.3.) does not necessarily imply that relevant health factors antedate the corresponding employment status. Indirect health selection implies that a common causal factor affects both health and employment status, but the sequence of relevant developments in health and employment status is not definite.

3.2. Aims of the present study

Data requirements for studying such complex dynamics are demanding. Cross-sectional surveys are able to describe the connections between health and employment status at a specific point in time, but are in principle unsuitable for investigating the processes which have led up to the observed pattern. Retrospective questions and the use of various control variables (cf., for instance, Arber et al. 1985; Bartley et

al. 1992; Søgaard 1993) are used in cross-sectional surveys in order to remedy these drawbacks, but inferences about dynamic relationships are nevertheless difficult to make.

Ideally, one needs data which continuously record every change in health situation and employment status for a large number of women, over a prolonged time period. Such data are rare – if they exist at all. Longitudinal data with repeated observations of the same individuals, i.e., panel designs, approach these requirements. The following study is based on surveys which, in addition to the cross-sectional parts, have panel data; both health and employment status are reported at more than one point in time, thus enabling the investigation of some of the dynamics described above. The aims of the following analyses are fivefold: (1) to describe how morbidity in terms of selected self-reported health indicators varied with women's main employment statuses in Norway during the 1980s, (2) to examine whether these differences were present at the point when women entered their first adult employment status, (3) to show how these differences were produced and reproduced by both changes in health and changes in employment status, (4) to shed light on the relative contribution of social causation and health-related mobility in generating these differences, and lastly (5) to judge whether the observed health-related mobility is direct, or indirect, health selection.

It should be noted that the concrete mechanisms of social causation or health-related mobility are not in focus here. Whether social support or self-actualization are the main causal factors within the social causation approach, or whether employers' discrimination or self-constraint are the key factors in health selection – to name but a few examples of such concrete mechanisms – will not be examined.

4. Data

The Norwegian Central Bureau of Statistics conducted four Surveys of Level of Living in 1980, 1983, 1987 and 1991 (Central Bureau of Statistics 1982, 1985, 1988, 1992), interviewing a sample generally representative of the population aged 16–79 years. Many questionnaire items were repeated unchanged in all four surveys. In 1983, part of the 1980 sample was reinterviewed, and this procedure was also followed in 1987 and 1991.

Accordingly, the surveys have three two-wave panels: 1980–83, 1983–87, and 1987–91, which may be pooled into a single two-wave panel data set in order to increase the sample size. This pooled two-wave panel sample has 1,822 female respondents aged 16–63 years at the time of the first interview – when women who reported being students either at the first or the second interviews are taken out of the sample, it is reduced to 1,604 respondents. It is also possible to construct a three-wave panel data set from these surveys (i.e., combining 1980–83–87 and 1983–87–91), but this sample is smaller and impractical to use, and therefore the following analyses will rely mainly on the pooled two-wave panel.

The three panels 1980–83, 1983–87 and 1987–91 are not independent of one another, as some respondents appear twice and some even thrice in the pooled data. The second time point observations in the 1980–83 panel are to some extent identical to the first time point observations of the 1983–87 panel, and so on. The panel data are in other words not freely sampled, and the variability is restricted. Furthermore, the panels are plagued, as most panel studies are, by sample attrition. Usually, the response rate is 76–79 per cent in these surveys (Central Bureau of Statistics 1982:14, 1985:14, 1988:14), but in the two-wave panel data, non-respondents accumulate, bringing the response rate down to about 66 per cent (1985:16, 1988:16). Accordingly, the representativeness of the pooled two-way panel sample is more questionable than usual in cross-sectional surveys.

Two health indicators are used in this article: limiting long-standing illness, and psychological problems. They are based on self-reporting: the respondents themselves classify and judge their health, and these indicators are not immediately congruent with judgements made by medical expertise. Both indicators are dichotomized here. Limiting long-standing illness is based on

a question posed to all respondents who had reported at least one long-standing illness, whether they felt that their illness limited their working capacity with respect to paid employment, housework, or school work, 'to some degree' or 'to a large degree'. The variable measuring psychological problems is based on three questions about experiences of palpitation, anxiety, and depression. Those who had at least one positive response are classified as having psychological problems. Being dichotomies, these indicators are of course not very sensitive, but they can be believed to divide the sample in a reasonable way into those who experience more, and those who experience fewer health problems.

Employment status is classified into four types: full-time employed (paid work 30 hours or more per week), part-time employed (1–29 hours per week), full-time homemakers, and a last category called 'pension'. Women without paid employment who reported that they did at least 500 hours unpaid housework the preceding year were classified as homemakers. The pension category consists mainly of recipients of public benefits such as disability pension, old age pension, rehabilitation support, long-term sickness benefits, unemployment benefits, benefits for single supporters, and so on. A few unclassifiable women were also placed in this category.

In addition to these four employment statuses, some respondents are pupils or students, who constitute the majority among those under the age of 20. The proportion of students falls rapidly thereafter, and very few are students among those aged 30+. Students are usually excluded from the analyses, as being a student must usually be considered as a status which precedes the employment statuses entered as adults. However, student respondents are analysed in order to describe the distribution of health among those who are recruited into 'the system' of employment statuses.

5. Results

5.1. Health inequalities among Norwegian women during the 1980s

The distribution of women's employment statuses changed considerably in the period studied (Table 1). More women became full-time employed (34 per cent in 1980, 45 per cent in 1991), more were classified in the pension category (8 per cent in 1980, 21 per cent in 1991), and full-time homemaking among women declined rapidly during the 1980s (28 percent in 1980, 10 per cent in 1991). These changes are interesting in themselves, but in this article the focus is on the inequalities in health displayed in the table. Table 1 shows that full-time employed women usually report somewhat fewer health problems than part-time employed. Homemakers come next – they experience more limiting long-standing illness and psychological problems than part-time employed. The pension category is dramatically more afflicted by health problems than the other three statuses.

In spite of large changes in the distribution of employment statuses, health differentials appear quite stable during these 11 years, and the distribution of ill-health in 1991 is not markedly different from the distribution in 1980. A few deviations from the main pattern can be observed: in 1983 a slightly higher proportion of limiting long-standing illness was reported among full-time employed women than among part-time employed women, and in 1987 homemakers reported fewer episodes of limiting illness than part-time employed women. These anomalies should probably be interpreted as instances of sample fluctuations. Thus, the pooled surveys displayed in the lower part of the table probably give the best estimates of the differences in self-reported morbidity during the 1980s. To combine these surveys (and to combine the corresponding panel data) has the obvious advantage of creating larger samples with more reliable estimates, and since there are no distinct time trends in differences in health during this period, this procedure seems legitimate.

Inequalities reported in these surveys are, by and large, similar to those found in many other studies from Western Europe and North America. It should be noted that the crude division of women into employed and non-employed used in some studies (for example Nathanson 1980; Verbrugge 1983) can readily conceal important differences within these categories. In Norway

in the 1980s, part-time employed women had somewhat less favourable health than full-time employed, and it is important to divide non-employed women into 'traditional' homemakers and the pension category. Homemakers have more health problems than employed women, but these differences are not dramatic, while the differences between homemakers and the pension category are very marked.

5.2. Health among entrants to adult employment statuses

As argued above, differences in health between employment statuses could be a reflection of women's initial health situation when they entered their first adult position as full-time or part-time employed, homemaker, etc. The surveys do not have direct information about this. However, they provide two approximations.

First, using the two-wave panel, we are able to sort out women aged 16–30 who were pupils or students at the first interview and who had entered an employment status at the second interview. From the two-wave panel sample, only 122 respondents can be included according to these criteria. Among those, only 15 were homemakers and seven were in the pension category at the second interview – therefore, these two categories of non-employed are combined. The interesting question concerns the distribution of health among these young women at the time when they still were students (Table 2, section A). Did their health, before they entered 'the world of adults', predict what employment status they eventually joined?

Second, the pooled cross-sectional surveys can be used in order to examine the connection between health and employment status among young women aged 16–25. The rationale behind this is that we would expect that health is not immediately influenced by a change of circumstances, i.e., by entering a particular employment status. Therefore, among women aged 25 or less, we would surmise that they have held their current employment status only for a relatively short period, and that their answers about current health problems reflect their health prior to their entries into the employment statuses of adult women (Table 2, section B).

To highlight possible contrasts between young and middle-aged women, the distribution of health among women aged 51–67 is also shown (Table 2, section C).

Table 2 indicates that when young women enter their first employment status, self-reported morbidity is not equally distributed among these statuses. The general pattern of inequalities is also found among entrants, in particular the favourable position of the full-time employed.

Thus, it is likely that inequalities in women's health associated with their employment status are already established from the start, so to speak, and this pattern is in some way reproduced among older women – note, for instance, that the differences in psychological problems are almost identical among women aged 16–25 and those aged 51–67.

With respect to the explanatory models discussed above, Table 2 strongly suggests that health-related mobility is involved in women's transition from their adolescent pupil/student status to their early adult employment statuses. In particular, the panel data (part A of Table 2) suggest that young women with few health problems tend to be recruited into full-time employment, while those who report more morbidity are some years later located in part-time employment, or become non-employed. However, this table cannot answer whether this health-related mobility is direct or indirect selection.

5.3. Stability or change in health and employment status

One reason for the correspondence between the general pattern of inequalities displayed in Table 1, and the pattern of inequalities among entrants displayed in Table 2A and 2B, could be that women tend to remain in the employment status they first entered, and that their health situation also tends to be relatively stable. Is stability in this sense responsible for the reproduction of the pattern of differences? Or are differences reproduced through changes, both in health and in employment status, throughout women's life courses?

Sections B and C in Table 2 show that there is a higher proportion of women aged 51–67 years afflicted by a limiting long-

standing illness than women aged 16–25 (44.0 per cent versus 18.4 per cent). Moreover, 630 out of 1,087 among the youngest respondents (58.0 per cent) as compared with 444 out of 1,830 (24.3 per cent) among the middle-aged are full-time employed. This suggests changes in employment status as women grow older, although it should be noted that age differences observed in cross-sectional data are not necessarily effects of age – they could be cohort effects (cf. Davies 1994).

However, these figures refer to marginal distributions, i.e., net changes. In order to answer the question about stability versus change in reproducing the differences, one needs information about gross changes, i.e., all individual moves between health categories and employment statuses. Such gross changes can be examined by constructing turnover tables from the two-wave panel sample. Table 3 indicates that although a majority stay in the same employment and health situation, a considerable number of changes occur. Only 44.3 per cent of homemakers and 56.5 per cent of part-time employed remained in the same employment status. Among those, 16.1 per cent who did not report any limiting long-standing illness at the first interview did so at the second, while 36.9 per cent among those who first reported such health problems were apparently free from them at the second interview. As a whole, it can be estimated from the table that 38.8 per cent moved to another employment status, 21.8 per cent reported a new situation regarding limiting long-standing illness, and 26.7 per cent experienced a change in psychological problems in terms of the measurements used here.

The frequency of employment status transitions is not surprising, considering the growth of women's participation in the labour market during the 1980s, and the well-known fact that women's paid employment tends to adjust to variations in family obligations. More unexpected were perhaps the numerous changes in state of health. That psychological problems vary considerably is not surprising, but it was unexpected that more than one-third of those who reported a long-standing illness that limited their working capacity at the first interview did not report such health problems three/four years later: it was assumed beforehand that a condition which is characterized as long-standing is likely to continue and not disappear some years later.

Studies of response reliability indicate a substantial amount of random variation in answers to health questions (Wärneryd 1991). Such measurement errors can lead to exaggerated notions of changes between panel waves. Whether long-standing illnesses should be classified as limiting or not limiting is sometimes an elusive question, and it is not surprising that the subjective evaluation involved produce fluctuations in the answers. But it is also clear that many long-term conditions improve over time, or vary as to their impact. Moreover, even if the seriousness of the condition has remained stable, its limiting effect on one's working capacity may be diminished if the woman has entered a new, less demanding, activity.

Therefore, it is not unrealistic that limiting long-standing illnesses reported in one interview are not reported some years later. Thus, although one may suspect that the number of changes in health problems are somewhat inflated in Table 3, an important and plausible finding is nevertheless that there are numerous movements and transitions, both in employment status and in health states. The frequency and profile of these changes and transitions are to some extent related to age, but it should be noted that all types of changes occur relatively often both among younger and middle-aged women (table not shown here).

Accordingly, it can be concluded that the link between employment status and ill-health observed at any specific point in time is not only a result of the distribution of ill-health among women when they first entered the 'system' of employment statuses, but also evolves through many changes and transitions in employment status as well as in health throughout the life-courses of women. The consequence of all this 'turbulence' is that the pattern of inequalities tends to be reproduced.

5.4. Social causation or health-related mobility?

The last section established that inequalities in health related to women's employment

Table 3. *Turnover tables (per cent). Women aged 16–63 at first interview. Norwegian Surveys of Level of Living, two-wave panel data.*

A: Employment status	Second interview					
First interview	Full-time	Part-time	Home-maker	Pension	Total	(N)
Full-time	72.7	15.3	6.1	5.9	100.0	(622)
Part-time	29.5	56.5	8.2	5.8	100.0	(549)
Homemaker	13.5	28.4	44.3	13.8	100.0	(327)
Pension	10.4	6.6	12.3	70.7	100.0	(106)
(N)	(669)	(505)	(241)	(189)		(1604)

B: Limiting long-standing illness	Second interview				C: Psychological problems	Second interview			
First int.	No	Yes	Total	(N)	First int.	No	Yes	Total	(N)
No	83.9	16.1	100.0	(1172)	No	79.1	20.9	100.0	(1016)
Yes	37.3	62.7	100.0	(432)	Yes	36.9	63.1	100.0	(588)
(N)	(1144)	(460)		(1604)	(N)	(1021)	(583)		(1604)

status are generated not only through stability (women stay on in their employment location and their health situation tends to remain the same), but also through numerous changes and transitions in these variables. In this section, the aim is to analyse further, by means of the two-wave panel data, how stability and change interact with each other. A particular purpose is to try to figure out how health and employment depend on each other – in other words, to examine whether social causation, or health-related mobility, or a mixture of these two models, is the dominant tendency. Given this two-wave panel sample, the general expectation is that health-related mobility is indicated by a data structure where employment status at the second interview is influenced by health situation at the first interview. In the case of social causation, the variables interact in the opposite manner: employment status at the first interview influences health at the second interview.

The two-wave panel data provide a cross-tabulation with four variables: employment status and self-reported health – either limiting long-standing illness or psychological problems – at the first interview, and the same variables observed three or four years later. This cross-tabulation has 64 cells, and log-linear analysis is a suitable method for analysing such a complex tabulation of categorical data (Knoke & Burke 1980; Norusis 1988; Hagenaars 1990). On the basis of theoretical reasoning, several unsaturated models will be formulated. The log-linear technique is used to determine the fit of these models by estimating the expected cell frequencies given that the model is correct, and by comparing the expected cell frequencies with those actually observed. Small deviations between expected and observed frequencies indicate a good model, i.e., a model which could have generated the actually observed frequencies. This is indicated by a relatively high probability of the likelihood ratio chi-square (L^2) compared to the corresponding degrees of freedom. This procedure determines whether or not the model generally agrees with the data, but a model may also be evaluated by comparing it to other models: one model is better than another if it results in a significant reduction of L^2 (Hagenaars 1990:56–68).

The first model (Model 1, Table 4) is based on three assumptions: (1) that the frequencies in the cross-tabulation are related to the marginal distributions of each variable; (2) that these frequencies are influenced by the association between

Table 4. *Relationships between health and employment status. Log-linear analyses, general frequency models. Women aged 16–63 at first interview, N = 1604. Two-wave panel data.*

Model	Test statistics			Reduction compared to Model 1		
	L^2	df	p-val	L^2	df	p-val
Health: limiting long-standing illness						
Model 1 (stability)	60.62	39	0.015			
Model 2A (health-related mobility, 1)	52.84	36	0.035	7.78	3	0.050
Model 2B (health-related mobility, 2)	36.00	27	0.115	24.62	12	0.018
Model 2C (health-related mobility, 3)	26.28	24	0.339	34.34	15	0.004
Model 3A (social causation, 1)	52.77	36	0.035	7.85	3	0.050
Model 3B (social causation, 2)	47.94	33	0.045	12.68	6	0.050
Model 3C (social causation, 3)	36.64	24	0.048	23.98	15	0.081
Model 4 (mixed)	44.70	33	0.084	15.92	6	0.015
Health: psychological problems						
Model 1 (stability)	39.83	39	0.433			
Model 2A (health-related mobility, 1)	33.89	36	0.569	5.94	3	0.122
Model 2B (health-related mobility, 2)	30.22	27	0.304	9.61	12	>0.250
Model 2C (health-related mobility, 3)	27.53	24	0.281	12.30	15	>0.250
Model 3A (social causation, 1)	35.35	36	0.500	4.48	3	>0.250
Model 3B (social causation, 2)	31.98	33	0.518	7.85	6	0.250
Model 3C (social causation, 3)	21.55	24	0.606	18.28	15	0.250
Model 4 (mixed)	26.78	33	0.769	13.05	6	0.044

employment status and health at both time points (thus, the terms HEA1*EMPL1 and HEA2*EMPL2 are included in the log-linear model – these terms express the findings in cross-sectional surveys that health and employment status are correlated); and (3) that an important tendency is stability, implying that a woman's employment status at the first interview tends to predict her employment status three/four years later, and that her health situation at the first time point influences her health situation when interviewed after three/four years (thus, the terms HEA1*HEA2 and EMPL1*EMPL2 are included in this model).

Model 2 assumes that in addition to the relationships included in the first model, the associations between health and employment status at the second interview are also generated through health-related mobility processes. If such processes are present, a woman's situation at the first interview will have an effect on what employment status she occupies later on. Model 2A reflects such possible effects by including a term which represents a direct link from the distribution of health at the first interview to the distribution of employment statuses at the second interview (HEA1*EMPL2). As it is possible that the influence of health on subsequent employment status can be mediated through interaction with other variables, two other models which represent health-related mobility are also fitted. Model 2B includes, in addition to Model 2A, the term HEA1*EMPL1*EMPL2, implying an interaction effect between health and employment status at the first interview on women's employment status at the second interview. Model 2C adds the term HEA1*HEA2*EMPL2, in other words assuming that the interaction between women's health at the first and second interviews influences what employment status she occupies at the second interview.

The third main model examines the social causation approach, which implies a causal chain from a woman's employment status

to her health or ill-health. Thus, this approach is represented by a model which assumes that the frequencies of the table are influenced, in addition to Model 1, by the association EMPL1*HEA2 (Model 3A). Analogous to the discussion about the second model, it is possible that the social causation influences are mediated through other variables. Therefore, Model 3B, which includes the three-variable interaction EMPL1*HEA1*HEA2, and Model 3C, which in addition to Model 3B includes EMPL1*EMPL2*HEA2, are also fitted.

Model 4 examines the combined two-variable effects of health-related mobility and social causation, by adding HEA1*EMPL2 and EMPL1*HEA2 to Model 1.

Concerning limiting long-standing illness, the stability model is insufficient ($L^2 = 60.62$, df = 39, p-val = .015). The 'best model' seems to be Model 2C, the health-related mobility model, version 3, which generates frequencies relatively close to the observed ones ($L^2 = 26.28$, df = 24, p-val = .339). This model indicates, however, that the effect of limiting long-standing illness on later employment status is not only a direct link, but also mediated through the three-variable relationships. In substantive terms, this suggests that the effect of limiting long-standing illness on a woman's employment status some years later varies both with her employment status at the first interview, and with her health development from first to second interview.

The analyses also indicate some social causation effects; all versions of Model 3 with limiting long-standing illness as the health variable have a significant or almost significant better fit than Model 1. Also, the mixed Model 4, which assumes a combination of health-related mobility and social causation effects, represents an improvement as compared with the stability Model 1 (reduction of $L^2 = 15.92$, df = 6, p-val = 0.015). But the effects associated with health-related mobility are clearly stronger than the social causation effects – compare, for instance, the test statistics for Model 2C and Model 3C.

As for psychological problems, the picture is somewhat different. Model 1 fits rather well ($L^2 = 39.83$, df = 39, p-val = 0.433). Neither health-related mobility models nor social causation models need to be included in order to account, in a statistically satisfactory way, for the frequencies of the cross-tabulation. This suggests that the correlation between psychological health and employment status is often generated by short-term adjustments, and that possible causal effects of one's employment status on psychological health, or vice versa, arise within a shorter time-span than can be examined using data which have observations three/four years apart. The drawback of not having continuous observations is evident here, and implies that the causal direction of short-term adjustments between employment status and psychological problems cannot be investigated. It should be noted, however, that Model 4 – the mixed model which combines Model 2A and Model 3A – appears to be the 'best model'.

The most reasonable interpretations of these log-linear analyses are perhaps the following: The associations between employment status and limiting long-standing illness are reproduced during a three/four-year period both by stability and by transitions and changes. These transitions and changes are primarily in accordance with the health-related mobility model, but the analyses also find developments which are compatible with social causation processes. Thus, we have a mixture of both types of processes, but health-related mobility clearly accounts for more of the transitions and changes than does social causation.

Concerning the association between employment status and psychological problems, the analyses also suggest a mixture of the two explanatory models, but neither strong causal effects of psychological health at the first interview on employment status four years later nor the opposite effects are found. Therefore, possible causal effects of psychological problems on employment status, or vice versa, are probably short-term and cannot be explored by these data which have observations three or four years apart. Thus, the analyses of psychological problems give no answer to the question about the relative contribution of the two explanatory models.

5.5. Health-related mobility: direct or indirect?

The preceding section found that health-related mobility contributes toward the generating of health inequalities related to women's employment statuses – in particular with respect to limiting long-standing illness. Health-related mobility may occur, however, in the form of direct selection – health is causally related to movements between employment statuses – or indirect selection – a woman's location in a particular employment status is primarily determined by other causal factors, and the health-related mobility effect arises because health correlates with these factors. The purpose of this section is to examine whether health-related mobility usually takes the form of direct, or indirect, selection.

An approach suitable for examining this question is to focus on employment status transitions and ask how these are predicted by health and by other possible causal factors. In the case of indirect health selection, the correlation between health and subsequent mobility will probably be spurious, because the causes of mobility are 'in reality' not health, but other causal factors. This means that indirect selection implies an association between health and subsequent mobility which diminishes, perhaps even disappears, when other control variables which more directly cause these employment status transitions are introduced in the analyses.

In this section, five types of employment status transitions from the first to second interview are analysed by means of logistic regression (Norusis 1990). These are movements from employment (full-time and part-time combined) to pension, from homemakers to employed and vice versa, and from full-time to part-time employed and vice versa – other types of transitions (cf. Table 3, part A) are not analysed because they involve only a small number of respondents.

Table 5 presents the effects of limiting long-standing illness and psychological problems measured at the first interview on the odds of making these five types of employment status transitions during the subsequent three/four years. The analyses are both bivariate and multivariate, thus allowing for a comparison of the effects of the health variable between the two designs.

Variables in addition to the health indicators are included because it is possible that they represent causal factors that have an effect on employment status transitions. It is hypothesized that middle-aged women have a higher probability of reducing or leaving paid employment, while younger women have a higher probability of increasing their participation in paid employment. Therefore, age is included, coded dichotomously into less than 45 at first interview (= 0) and 45 + (= 1). Education (measured by years) is assumed to influence mobility in an opposite manner – more education may increase the probability of paid employment. Furthermore, change in family situation can be influential. Logically, such changes should have been measured prior to changes in employment status. With a two-wave panel this is not possible, and instead, the analyses include variables which indicate changes in family situation between the first and second interviews: change in the number of children aged 0–7 years in the household (Ch. childr.), and change in marital status (Ch. marital). Thus, respondents are divided into stability (= 0) versus change (= 1) with regard to family situation. Lastly, it is assumed that labour-market characteristics influence women's mobility. Opportunities for paid work are believed to differ between rural and urban districts, and to change over time. These aspects are indicated by the variable 'Urban' (place of living at first interview, coded rural = 0, urban = 1) and by the dummies 1980–83 and 1983–87, which indicate variations between the three panel periods – the last period, 1987–91, is the reference category.

Table 5 suggests many processes which influenced women's movements between employment statuses during the 1980s. Age and education have, as hypothesized, opposite effects. Change in the number of small children during the period is connected with change in employment status during the same period, while change in marital status has practically no effect. The rural–urban dimension makes almost no difference

Table 5. *Effects of limiting long-standing illness, psychological problems, and other variables on employment status transitions. Logistic regression. Women aged 16–63 at first interview. Two-wave panel data.*

From	Employed		Homemaker		Employed		Full-t. empl.		Part.-t. empl.	
To	Pension		Employed		Homemaker		Part-t. empl.		Full-t. empl.	
Movers (N)	69		137		83		95		162	
Stayers (N)	1019		145		1019		452		310	
Odds (M/S)	0.068		0.945		0.081		0.210		0.523	
	B	p-val	B	p-val	B	p-val	B	p-val	B	p-val
Bivariate										
Limit. l-s illn.	0.860	0.001	−0.200	0.476	0.460	0.067	0.553	0.036	0.110	0.623
Psych. problems	0.803	0.001	−0.326	0.185	0.516	0.027	0.526	0.029	0.002	0.991
Multivariate										
Limit. l-s illn.	0.859	0.002	−0.012	0.973	0.411	0.119	0.487	0.083	0.214	0.363
Age	1.217	0.000	−0.439	0.167	0.478	0.118	0.116	0.677	−0.518	0.026
Y of Educ.	−0.287	0.000	0.140	0.068	−0.183	0.004	−0.137	0.015	0.094	0.060
Ch. childr.	0.720	0.063	0.438	0.141	1.632	0.000	1.066	0.000	−0.525	0.046
Ch. marital	0.163	0.695	0.721	0.205	0.211	0.554	0.529	0.091	0.404	0.333
Urban	0.171	0.574	0.102	0.709	0.168	0.549	−0.747	0.005	0.184	0.438
1980–83	−1.332	0.000	−0.555	0.145	0.312	0.272	0.436	0.138	−1.000	0.000
1983–87	−1.336	0.000	−0.038	0.922	−0.235	0.438	0.263	0.364	−0.478	0.042
Psych. problems	0.699	0.008	−0.306	0.243	0.409	0.090	0.487	0.053	−0.015	0.943
Age	1.270	0.000	−0.452	0.151	0.507	0.098	0.165	0.554	−0.504	0.030
Y of Educ.	−0.275	0.000	0.126	0.102	−0.178	0.005	−0.128	0.022	0.091	0.069
Ch. childr.	0.708	0.068	0.450	0.131	1.622	0.000	1.077	0.000	−0.535	0.043
Ch. marital	0.145	0.726	0.751	0.188	0.199	0.577	0.523	0.096	0.425	0.310
Urban	0.122	0.686	0.128	0.639	0.132	0.636	−0.781	0.003	0.191	0.424
1980–83	−1.201	0.000	−0.554	0.147	0.344	0.222	0.452	0.124	−0.991	0.000
1983–87	−1.284	0.000	−0.040	0.919	−0.231	0.446	0.305	0.291	−0.478	0.041

(except concerning transitions from full-time to part-time employment, which occurs more frequently in rural districts), while the coefficients for the period variables 1980–83 and 1983–87 indicate that the probability of transitions from employed to pension and from part-time to full-time was higher during the late 1980s than earlier.

In this study, however, the effects of the health indicators are in focus. The bivariate analyses show that both limiting long-standing illness and psychological problems have a statistically significant effect on the probability of three types of transition: from employment to pension, from employment to homemaking, and from full-time to part-time employment. Movements which imply increased participation in paid employment (from homemaking to employed, or from part-time to full-time), are only insignificantly affected by these health variables, however. Thus, it seems that health-related mobility generates inequalities primarily through transitions which imply a reduction in paid employment – changes in employment status away, so to speak, from full-time employment.

On the question of direct or indirect selection, differences between the bivariate and the multivariate analyses are relevant. With respect to transitions from employment to pension, B-values for limiting long-standing illness are 0.860 bivariately and 0.859 multivariately, and the corresponding B-values for psychological problems are 0.803 and 0.699. Thus, the coefficients are virtually of the same size. It seems that the association is not spurious, and this indicates a process of direct health selection. This is not surprising, because this type of transition is often a state-regulated health selection (described in section 2.2.),

and it is not unexpected that health is involved here in a manifest way.

But neither are the effects of health much reduced from the bivariate to the multivariate analyses with regard to the other two relevant types of transition – from employment to homemaking, and from full-time to part-time employment. This is striking, and it means that when women make such changes in their employment status, it is likely that there are evident health reasons for them – in combination with other causal factors, as the analyses show that, in addition to health problems, low education and change in the number of small children predict transitions from employment to homemaking and from full-time to part-time employment.

These findings indicate that, above all, the health-related mobility observed in these data seems to be of the direct selection type. The analyses support the hypothesis that health is causally related to later employment status transitions. We cannot reject the existence of indirect selection, of course, as it is uncertain whether there are other important factors which are not controlled for (such as social background). Nevertheless, a reasonable interpretation of Table 5 is that health-related mobility found in these data primarily means reductions in paid employment with health problems as a directly contributing causal factor.

6. Conclusion

This study has shown that there existed inequalities connected with employment status in women's self-reported health in Norway during the 1980s, by and large similar to inequalities found in many other studies. Employed women, in particular full-time employed, reported relatively few health problems, while non-employed generally, and especially women receiving some kind of public benefits, reported more health problems.

The main purpose of this study was to describe and analyse the dynamics behind these associations. In particular, the study aimed at examining whether these dynamics suggest that the pattern of inequalities is generated through social causation or health-related mobility, or both. In general, the results indicate that both health-related mobility and social causation are contributory factors in generating these health differentials, but health-related mobility appears to be the most significant of these two explanatory models – contrary to what seems to be the prevailing view.

Health inequalities are already present when women enter their first employment status. Thus, selective processes are involved in women's transitions from adolescence to adulthood, which tend to recruit young women with relatively few health problems into full-time employment, while those who report having health problems tend to enter part-time employment or to remain non-employed.

Moreover, the pattern of inequalities – already in existence among young women – is produced and reproduced through frequent changes of employment status and frequent improvements and deteriorations in health during women's life-courses. In this study, these dynamics have been analysed by examining how the associations between health and employment status evolve during a period of three/four years. The analyses of these dynamics indicate both social causation and health-related mobility processes, but the pattern is usually more in accord with the health-related mobility explanation – at least this is clearly evident with regard to limiting long-standing illness. Analyses of the dynamics behind the generation of inequalities are hampered, however, by the data structure, which does not allow for analyses of short-term adjustments between health and employment status.

Not all types of employment status transitions are equally related to health, however. These data indicate that health-related mobility primarily applies to those employment status transitions which imply reductions in paid employment and movements out of the labour market. The opposite process – for instance homemakers who enter paid employment – seems generally unrelated to health. Thus, inequalities are generated by mobility 'downwards', so to speak – by transitions which locate women with health problems in employment

statuses other than full-time employment – into part-time employment, homemaking, and in particular into the pension category. Moreover, the data suggest that this health-related mobility occurs primarily in terms of direct health selection – the analysis of these transitions indicates that health characteristics are involved in a direct, manifest way. Thus, health problems seem causally related to these movements, and indirect selection is not found in these analyses.

A main conclusion of this study is that health-related mobility perhaps plays a more prominent part in generating inequalities in health associated with women's employment status than has usually been believed. It could be that the special historical circumstances have contributed to these results: the 1980s in Norway was a turbulent period in women's relationship to paid employment and other employment statuses, and perhaps the role of health-related mobility was particularly enhanced by this situation. Therefore, whether or not these results can be generalized to other social contexts is unclear.

Inequalities in women's health connected to their employment statuses are produced and reproduced through complex dynamic processes. This study was based on a two-wave panel data set, which allows for several types of analyses which cross-sectional data cannot handle. Nevertheless, the opportunities for modelling the processes have been restricted. Data cover a relatively short time-span, because respondents are followed only for three or four years, and it is problematic that data only have observations at two points in time – when a respondent experiences changes both in health and in employment status between the first and second interviews, the data do not permit any time ordering of these transitions. These drawbacks should of course be borne in mind when evaluating the results of this study.

Two directions for further research on these topics are suggested by this study. First, if panel data covering a longer time period with more frequent recording of changes in women's health and employment status were available, more precise analyses of these dynamics would be possible, and judgements of the relative contribution of social causation and health-related mobility could be more precise. Second, this study indicates that health-related mobility plays an important part in generating these inequalities, in particular in terms of employment status transitions away from full-time employment caused, wholly or partly, by health problems. A follow-up of these findings could be a more detailed investigation of the mechanisms involved, in order to determine the part played by state-regulated health selection, employer discrimination, and other possible factors in producing this pattern.

Acknowledgements

The data material used in this article was supplied by the Norwegian Social Science Data Services. Data were originally collected and processed by the Central Bureau of Statistics, Norway. Neither of these institutions is responsible for the data analyses or the interpretations provided here. The author thanks two anonymous reviewers for constructive criticism of an earlier version of this article.

First version received December 1994
Final version accepted June 1995

References

Arber, S. 1990. Opening the 'Black Box': Inequalities in Women's Health. In P. Abbott & G. Payne (eds.), *New Directions in the Sociology of Health*. London, New York and Philadelphia: The Falmer Press.

Arber, S. 1991. Class, Paid Employment and Family Roles: Making Sense of Structural Disadvantage, Gender and Health Status. *Social Science & Medicine*, 32, 425–436.

Arber, S., Gilbert, G. N. & Dale, A. 1985. Paid Employment and Women's Health: A Benefit or a Source of Role Strain? *Sociology of Health & Illness*, 7, 375–400.

Arber, S. & Lahelma, E. 1993a. Inequalities in Women's and Men's Ill-health: Britain and Finland Compared. *Social Science & Medicine*, 37, 1055–1068.

Arber, S. & Lahelma, E. 1993b. Women, Paid Employment and Ill-health in Britain and Finland. *Acta Sociologica*, 36, 121–138.

Axelsson, C. 1992. *Hemmafrun som försvann. Övergången til lönearbete bland gifta kvinnor i Sverige 1968–1981* [The Housewife that Disappeared. Married Women's Transition to Paid Employment in Sweden 1968–1981]. Stockholm: Institutet för social forskning, Avhandlingserie 21.

Barker, D. J. P. 1991. The Foetal and Infant Origin of Inequalities in Health in Britain. *Journal of Public Health Medicine*, 13, 64–68.

Bartley, M. 1991. Health and Labour Force Participation: 'Stress', Selection and the Reproduction Costs of Labour Power. *Journal of Social Policy*, 20, 327–364.

Bartley, M., Popay J. & Plewis, I. 1992. Domestic Conditions, Paid Employment and Women's Experience of Ill-health. *Sociology of Health & Illness*, 14, 313–343.

Blane, D., Smith, G. D. & Bartley, M. 1993. Social Selection: What Does it Contribute to Social Class Differences in Health? *Sociology of Health & Illness*, 15, 1–15.

Bloor, M., Samphier, M. & Prior, L. 1987. Artefact Explanations of Inequalities in Health: An Assessment of the Evidence. *Sociology of Health & Illness*, 9, 231–264.

Central Bureau of Statistics, Norway. 1982. Survey of Level of Living 1980. *NOS B 320*, Oslo, Kongsvinger.

Central Bureau of Statistics, Norway. 1985. Survey of Level of Living 1983. *NOS B 511*, Oslo, Kongsvinger.

Central Bureau of Statistics, Norway. 1988. Survey of Level of Living 1987. *NOS B 772*, Oslo, Kongsvinger.

Central Bureau of Statistics, Norway. 1992. Survey of Level of Living 1991. *NOS C 43*, Oslo, Kongsvinger.

Central Bureau of Statistics, Norway. 1993. Sosialt utsyn 1993 [Social Survey 1993], Oslo, Kongsvinger.

Dahl, E. 1994a. Sosial ulikhet i helse: Artefakter eller seleksjon? [Social Inequalities in Health: Artefacts or Health Selection?]. Doctoral Thesis. *FAFO-Rapport 170*. Oslo: Institute of Applied Social Science (FAFO).

Dahl, E. 1994b. Health Inequalities: What's Left of the Health Selection Explanation? *Draft*. Oslo: Institute of Applied Social Science (FAFO).

Davies, R. B. 1994. From Cross-sectional to Longitudinal Analysis. In A. Dale & R. B. Davies (eds.), *Analyzing Social and Political Change. A Casebook of Methods*, pp. 20–40. London, Thousand Oaks and New Delhi: SAGE Publications.

Elstad, J. I. Forthcoming. Inequalities in Women's Health Related to Women's Marital, Parental, and Employment Status – A Comparison Between the Early 70s and Late 80s, Norway. *Social Science & Medicine*.

Forsdahl, A. 1977. Are Poor Living Conditions in Childhood and Adolescence an Important Risk Factor for Arteriosclerotic Heart Disease? *British Journal of Preventive and Social Medicine*, 31, 91–95.

Fox, A. J., Goldblatt, P. O. & Jones, D. R. 1986. Social Class Mortality Differentials: Artefact, Selection, or Life Circumstances? *In* R. G. Wilkinson (ed.), *Class and Health. Research and Longitudinal Data*, pp. 34–49. London: Tavistock.

Gerhardt, U. 1989. *Ideas About Illness. An Intellectual and Political History of Medical Sociology*. Houndmills and London: Macmillan.

Gove, W. R. 1984. Gender Differences in Mental and Physical Illness: The Effects of Fixed Roles and Nurturant Roles. *Social Science & Medicine*, 19, 77–91.

Gove, W. R. & Geerken, M. 1977. The Effects of Children and Employment on the Mental Health of Married Men and Women. *Social Forces*, 56, 67–76.

Hagenaars, J. A. 1990. *Categorical Longitudinal Data. Log-linear Panel, Trend, and Cohort Analysis*. Newbury Park, London, New Delhi: SAGE Publications.

Hibbard, J. H. & Pope, C. R. 1987. Employment Characteristics and Health Status Among Men and Women. *Women & Health*, 12, 85–102.

Hibbard, J. H. & Pope, C. R. 1991. Effect of Domestic and Occupational Roles on Morbidity and Mortality. *Social Science & Medicine*, 32, 805–811.

Illsley, R. 1986. Occupational Class, Selection and the Production of Inequalities in Health. *The Quarterly Journal of Social Affairs*, 2, 151–165.

Knoke, D. & Burke, J. 1980. *Log-linear Models*. Beverly Hills and London: SAGE Publications.

Lundberg, O. 1990. *Den ojämlika ohälsan. Om klass- och könsskillnader i sjuklighet.* [Inequality in Ill-health. On Class and Sex Differences in Illness.] Stockholm: Institutet för social forskning.

Lundberg, O. 1991. Childhood Living Conditions, Health Status, and Social Mobility: A Contribution to the Health Selection Debate. *European Sociological Review*, 7, 149–162.

Nathanson, C. 1980. Social Roles and Health Status Among Women: The Significance of Employment. *Social Science & Medicine*, 14a, 463–471.

Norusis, M. J. 1988. *SPSS-X Advanced Statistics Guide*. Chicago: SPSS Inc.

Norusis, M. J. 1990. *SPSS/PC+ Advanced Statistics 4.0*. Chicago: SPSS Inc.

Næss, S. 1986. Yrkeskvinne – husmor? Gifte kvinners livskvalitet. [Employed Woman – Housewife? The Quality of Life of Married Women.] *INAS-Report 86:3*. Oslo: Institute of Applied Social Research.

Passannante, M. R. & Nathanson, C. 1985. Female Labor-force Participation and Female

Mortality in Wisconsin 1974–78. *Social Science & Medicine*, 21, 655–665.
Rosenfield, S. 1989. The Effects of Women's Employment, Personal Control and Sex Differences in Mental Health. *Journal of Health and Social Behavior*, 30, 77–91.
Smith, K. R. & Waitzman, N. J. 1994. Women, Work, and Whether Occupation Matters: Differences in Mortality by Occupation in the U.S. 1971–1987. Research Committee 19, XIII World Congress of Sociology, Bielefeld, Germany.
Sorensen, G. & Verbrugge, L. M. 1987. Women, Work and Health. *Annual Review of Public Health*, 7, 235–251.
Statistics Norway. 1995. Historical Statistics 1994. Oslo, Kongsvinger: *Official Statistics of Norway C 188*.
Sogaard, A. J. 1993. Er det egentlig sunt å jobbe? [Is Work Really Healthy?]. In B. Schei, G. Botten & J. Sundby (eds.), *Kvinnemedisin* [Women's Medicine], pp. 221–35. Oslo: Ad Notam-Gyldendal.
Thebaud-Mony, A. 1994. The Production of the Social Inequalities in Health. *Paper*, 5th Biennial Conference of European Society of Medical Sociology, Vienna, 16–18 September.
Townsend, P. & Davidson, N. 1982. *Inequalities in Health. The Black Report*. Harmondsworth: Penguin.
Verbrugge, L. M. 1983. Multiple Roles and Physical Health of Women and Men. *Journal of Health and Social Behavior*, 24, 16–30.
Verbrugge, L. M. 1986. Role Burdens and Physical Health of Women and Men. *Women & Health*, 11, 47–77.
Vågerö, D. 1994. Women, Work and Health – Has the Increasing Female Labour Participation Contributed to Women's Improved Health, or Not? An Analysis of Female Mortality Rates in Sweden 1981–86. *Paper*, Research Committee 19, XIII World Congress of Sociology, Bielefeld, Germany.
Waldron, I. 1980. Employment and Women's Health: An Analysis of Causal Relationships. *International Journal of Health Services*, 10, 435–54.
Waldron, I. & Jacobs, J. A. 1989. Effects of Multiple Roles on Women's Health – Evidence From a National Longitudinal Study. *Women & Health*, 15, 3–19.
Wärneryd, B. 1991. *Levnadsförhållanden, Appendix 12. Återintervjustudie i undersökningen av levnadsförhållanden (ULF) 1989* [Conditions of Living, Appendix 12. A Study of Reinterviews of Living Conditions 1989.] Stockholm: Statistiska centralbyrån.
West, P. 1991. Rethinking the Health Selection Explanation for Health Inequalities. *Social Science & Medicine*, 32, 373–384.
Wilkinson, R.G. 1986. Socio-economic Differences in Mortality: Interpreting the Data on Their Size and Trends. *In* R. G. Wilkinson (ed.), *Class and Health. Research and Longitudinal Data*, pp. 1–20. London: Tavistock.

Paper V:

"Health-related mobility, health inequalities and gradient constraint: discussion and results from a Norwegian study", *European Journal of Public Health*, 2001, forthcoming.

Health-related mobility, health inequalities and gradient constraint: discussion and results from a Norwegian study

Jon Ivar Elstad
NOVA – Norwegian Social Research
Box 3223 Elisenberg
N-0208 Oslo
Norway

Telephone: (+47) 22 54 12 88 / 22 54 12 00
Fax: (+47) 22 54 12 01
E-mail: Jon.I.Elstad@isaf.no

Revised version of Manuscript 9B 1187,
cf. comments from language editor,
April 2000

Word count of body text (excluding abstract, references, tables):
3,460 words.
Three tables.

Running head:
Mobility and health inequalities

Abstract

Background. Recent studies have argued that health-related mobility does not widen social class health differentials, but rather moderates them. This is termed gradient constraint. This paper examines gradient constraint from a theoretical and empirical angle.

Methods. How health-related mobility influences social class health differentials is discussed using hypothetical models. In a Norwegian survey with data on intergenerational mobility (N = 1,853 males aged 30-69 years), mean health and height values for different subsections of the sample were analysed.

Results. When initial social class health differences are large and mobility widespread, health-related mobility may lead to smaller differentials, but the result depends on how closely mobility varies with health. This empirical study found non-significant increases in height and health differentials from class of origin to class of destination. The interpretation has to consider effects of both social mobility and social causation. Health is measured in the post-mobility situation and the effects of social mobility and social causation are difficult to separate from each other for most of the health indicators analysed. However, this limitation does not apply to height which is not influenced by social causation during adulthood. In these data mobility did not reduce the height differential.

Conclusion. Health-related mobility can either lead to smaller or larger social class health differentials. The specific effects of social mobility cannot be determined without knowing how social causation has interfered. The intergenerational mobility process analysed in this paper does not show gradient constraint as regards the height differential between the worker and higher occupational categories.

Keywords: gradient constraint, health inequalities, health-related mobility, social causation, social class

INTRODUCTION

Several studies have indicated that upward mobility is associated with good health and downward mobility with poor health.[1-9] A disputed question is how much such health-related mobility contributes to the inverse relationship between health and socioeconomic status. Some[3,10] have suggested that mobility is important, while others have contended that health-related mobility is of little significance compared to social causation.[11-20]

Recently, studies[21-23] have advanced new arguments against the hypothesis that social inequalities in health are produced by social mobility. Their findings indicate that mobility does not widen social class health differences, but rather constrains them. This is termed gradient constraint: 'Gradient constraint can be described as the process whereby social mobility moderates, rather than creates or widens, the size of the social class differential' (p.68).[22]

These studies addressed both intragenerational mobility[21,22] and intergenerational mobility.[23] They found that mobility is health related: those who have experienced upward mobility have less limiting long-standing illness,[21] lower mortality,[22] and higher stature[23] than those who remain in the lower classes and vice versa among those who have experienced downward mobility. However, the main focus was on in-flow mobility, i.e. the stage in the mobility process where the 'new' classes are formed by the stable members of the class and those who move into these destination classes. Here they observed a characteristic pattern. Upward movers into the upper classes have worse health than stable members of the upper classes, while downward movers have better health than stable members of the lower classes. It is argued that this is the reason why mobility tends to moderate health differentials. As upward movers are less healthy and downward movers healthier than the stable members of the destination classes, mobility tends to 'dilute the levels of morbidity in each class' (p.377)[21] and therefore constrain the magnitude of health differentials.

Gradient constraint, i.e. the mobility process which moderates the size of the social class health differentials, is proposed as the reason why many longitudinal studies (including Bartley and Plewis[21]) have not found increasing health differences over time. This is contrary to what might be expected from the usual explanations of health inequalities. If health differences are caused by more health-damaging environments in lower than in upper classes (social causation), one would expect that length of exposure would lead to a widening differential. If there is a constant sorting of individuals into socioeconomic

status groups according to their health (health-related mobility), gradients could also be expected to widen as mobility continuously directs people with good health upwards and people with bad health downwards in the social hierarchy.[21,22] It is suggested that social causation leads to wider health gradients while social mobility acts in the opposite way: 'The size of social class differences in health may remain comparatively stable through life because they are the net effect of disease-producing processes, which over time tend to widen the differences, counterbalanced by gradient constraint, which acts in the opposite direction' (p.68-69).[22]

The concept of gradient constraint is interesting as it contradicts the widely held belief that health-related mobility always contributes to the production of health gradients. In this paper this concept will be investigated further from both a theoretical and an empirical angle.

First, using models with hypothetical data, we will show how gradient constraint, i.e. mobility which reduces the size of social class health differentials, may emerge. We will also show how mobility may lead to the opposite outcome. Thus, both smaller and larger differentials may emerge as a consequence of health-related mobility. Some reasons for this are discussed. Next, in the empirical section, we will analyse Norwegian data on intergenerational mobility and health and discuss whether gradient constraint is present in these data.

HOW HEALTH-RELATED MOBILITY MAY INFLUENCE HEALTH GRADIENTS

In this section, hypothetical models are used to discuss the mechanisms which are involved in how health-related mobility affects the size of health differentials.

How health-related mobility may reduce the health differential is illustrated by example A in *table 1*. For simplicity, we assume only two social classes and two health statuses and that the health of every individual is constant. The initial health differential is 20%. Mobility is health related: the probability of downward mobility from the upper class is 0.4 if health is bad and 0.3 if health is good. Similar differences in mobility probabilities exist in the lower class.

In the mobility process, movers from the upper class (120 with good health and 40 with bad health) join the stable members of the lower class. In the same manner the 'new' upper class is formed by joining 340 stable members of the upper class with 360 upward movers from the lower class. As a consequence, the proportion of bad health in the lower class is constant while bad health in the upper class increases. The differential is reduced from 20 to 14%.

Example A demonstrates gradient constraint: health-related mobility reduces the size of the health differential. This is of course a hypothetical result, but it is suggestive that the model is constructed so that the data resemble what we often find in reality: a considerable health difference at the beginning, widespread mobility and some but not particularly large differences in mobility probabilities between those with good and bad health. Given such circumstances, gradient constraint is not unlikely. An intuitive interpretation is that, although mobility is associated with health, it is also associated with other factors such as educational opportunities and changes in the occupational structure. When mobility is widespread, not only many 'healthy' people but also a number of 'sick' people from the lower class move upwards and some 'healthy' people move downwards and this interchange leads to a smaller health differential.

As example A has many similarities with real conditions, it is not unreasonable to believe that gradient constraint, i.e. mobility which moderates the size of health differentials, perhaps occurs quite often in our societies. However, the result displayed in example A is due to the particular circumstances pertaining to this example and, if these circumstances are changed, the result may easily be otherwise.

This is illustrated by example B. This example is identical to example A except for the mobility probabilities. The probability of moving down from the upper class if health is good is reduced from 0.3 to 0.2, and the same reduction in the probability of moving up from the lower class if health is bad is introduced. Thus, the selectivity for health in the mobility transitions is stronger, i.e. mobility covaries more closely with health. Now, the health differential increases from 20 to 25%. A seemingly small change in the selectivity for health leads to quite diverging results. A relatively weak selectivity for health leads to a reduced health differential (example A), while the stronger selectivity for health in example B results in a larger health differential.

However, selectivity for health is only one of the circumstances which determine the result. Using models similar to those presented in *table 1*, it can be shown that the outcome will also depend on the size of the initial health differential (health-related mobility tends to reduce the health differential if the initial differential is large and to increase the differential if the initial differential is small) and on the overall rate of mobility (when mobility is widespread, its potential for changing the health differentials is greater).

The earlier studies[21-23] have suggested that gradient constraint occurs when upward movers are less healthy than stable members of the upper class and

downward movers have better health than stable members of the lower class. However, examples A and B show that this particular in-flow mobility pattern is not in itself sufficient to produce gradient constraint. This pattern is present in both example A where the differential is reduced and in example B where the differential is increased. More bad health among upward movers and better health among downward movers in comparison with the stable members is therefore compatible both with reduced health differentials (gradient constraint) and with wider health differentials. The specific focus in the earlier studies on the in-flow health differentials is somewhat misleading as the outcome is determined by the whole mobility process, not only by the health differences between stable members and movers into the destination classes, but also by the health of leavers and, most importantly, by the proportion of the sample which belongs to these different categories.

In conclusion, as proposed by the earlier studies, gradient constraint, i.e. a reduced health differential as a consequence of mobility, may typically occur when the selectivity for health in the mobility transitions is relatively weak, the overall rate of mobility is high and initial health differentials are considerable. However, health-related mobility may also result in wider health differentials and whether gradient constraint has occurred or not in a particular mobility process must be studied empirically. We will therefore next analyse a Norwegian survey with data on intergenerational mobility and health and discuss whether gradient constraint is found in these data. We cannot neglect the possible role of social causation with real data, as was done in the hypothetical models. How this complicates the interpretation will be discussed below.

DATA AND METHODS

In 1995, the Norwegian Health Survey interviewed respondents from a nationally random sample.[24] Information was collected by personal interviews (response rate 77.8%) and by a questionnaire. Of those interviewed 10.4% did not return the questionnaire. The resulting overall response rate is close to 70%. Men aged 30-69 years are analysed. Among younger respondents own class of destination was often not fixed[23] and among older respondents differential mortality could have distorted the representativity of the sample.

Own occupational class is coded according to the standard classification of Statistics Norway: unskilled/skilled workers, lower/medium/higher white collar, farmers and fishers and other self-employed.[25] One thousand, six hundred and sixty-one respondents were classified according to own current occupation, while 356 respondents were without employment at the time of the interview (or

information was missing). The majority of the non-employed received an old age pension or, more frequently, disability pension because of health problems. As disability pensions disproportionally often affect men in low-status occupations,[26] the picture of health inequalities will be biased if based only on those who are currently employed. The survey also asked about main previous occupation. In addition 301 non-employed respondents could be classified. Information about own (current or previous) occupational class was obtained for 1,962 respondents.

The respondents were furthermore asked about their fathers' occupation using pre-structured response alternatives corresponding to the occupational categories described above. One thousand, nine hundred and three respondents gave information about their fathers' occupation. One thousand, eight hundred and fifty-three respondents (92% of male respondents aged 30-69 years in the survey) with information about both their own and fathers' occupational class are analysed here.

The occupational classification was collapsed into three categories: workers (unskilled plus skilled), intermediate occupations (lower white collar, farmers/fishers and other self-employed) and higher occupations (medium and higher white collar). This was partly done because a number of cells in the full mobility table had very few respondents. Moreover, the dividing line between unskilled and skilled workers and between medium and higher white collar occupations is not clear.[21] To decide what constitutes upward and downward mobility is further complicated because own and father's occupational class refer to very different economic circumstances.

Using this collapsed classification, moves between the worker and higher category can be considered as vertical intergenerational mobility. Movements concerning the intermediate occupations are more ambiguous. A few farmers and self-employed people are wealthy, but many in the intermediate category are smallholders and minor shopkeepers. Their location 'above' workers can be questioned. As the data do not permit a more precise classification, it is reasonable to place these categories as intermediate.

Five variables indicate the health of respondents. Whether height is an indicator of health is debated. It is included here because it was used in one of the earlier studies[23] and has often been found to be associated with mobility and health.[3,5,12] Self-perceived health is based on a five-item question. The variable shows the proportion reporting self-perceived health below good. Somatic symptoms were measured by ten questions about back and shoulder pain,

digestive problems, etc. The respondent was asked whether he had recently been troubled 'not at all, a little, moderately or very much' by these symptoms. The variable counts the number of moderately/very much responses. Similarly, the variable mental symptoms counts the number of moderately/very much responses to the 25-item Hopkins Symptom Check List.[27,28] Lastly, the variable medical diagnoses counts the number of medical diagnoses reported.

The data do not show when respondents' health conditions emerged. Respondents were asked about their current situation. Their reported health problems may have preceded mobility transitions, but may also have emerged after respondents had entered their final class of destination. The implications of this will be addressed later. However, this limitation does not apply to height, which is established in adolescence before their adult occupational career.

In the following, the focus is on the average health values for different sections of the sample. The size of the health differentials and how they change are analysed. In order to adjust for possible confounding effects of age compositions, all subsamples have been directly standardised for age, according to the age distribution (ten-year bands) of the whole sample. Confidence intervals are included for results which are of particular interest for the arguments, i.e. the health differentials for class of origin and class of destination and the change in these health differentials.

RESULTS

Table 2 indicates substantial intergenerational social mobility. Less than half of the respondents (787 out of 1,853, 42.5%) were in the same occupational classes as their fathers (the cells along the diagonal), 39.7% had experienced upward mobility as defined here (the cells to the right of the diagonal) and 17.8% have been mobile downwardly.

Average health is shown according to class of origin and class of destination in *table 3*. Those with a working-class background were, on average, shorter and reported more health problems than those with fathers in the higher occupations. In addition, when classified according to own class (class of destination), we observed the health disadvantages of workers. The values for the intermediate occupations were usually (but not always) between workers and higher occupations. These irregularities were probably due to the difficulties in locating the intermediate occupations precisely in the social hierarchy.

Table 3 also focuses on the differentials between workers and higher occupations. A small increase in height and health differentials was observed from class of origin to class of destination for all health indicators. However, these increases were not significant in statistical terms (the lower limit of the confidence intervals for the changes was always below zero).

The changes in differentials from class of origin to class of destination arose as a result of out-flow and in-flow mobility. *Table 3* also shows how this occurred for the worker and higher occupational categories. A pattern indicating health-related out-flow mobility can be seen. Those who had left the working class and were upwardly mobile, either towards the intermediate or towards the higher occupational category, were of higher stature and had better health values than those who remained. Among those who had fathers in higher occupations, leavers had less favourable health values than stable members. Thus, out-flow mobility 'subtracts' a relatively healthy group from the working class and a relatively unhealthy group from the higher occupational category.

Next, in-flow mobility forms the destination classes. Usually, those who were downward movers had better health and upward movers less good health than stable members of the classes they moved into. Movers to the working class, who were most often from an intermediate background (*table 2*), were in better health than the stable working class respondents with the exception of average height. Entrants upwards from the working class and the intermediate category into the higher occupations were of shorter stature and worse health than stayers in this upper part of the hierarchy (except for medical diagnoses, where the values were practically equal).

DISCUSSION

This study suggests that intergenerational social mobility in this particular sample of Norwegian men is associated with health as upward movers have better health values and downward movers less favourable health values than the stable members of the classes they leave. However, the main interest in this study was whether mobility had acted to reduce the health differentials, i.e. whether gradient constraint was present in these data.

To answer this question, we need to estimate the health differentials before and after mobility. The class of destination health differentials indicate the post-mobility differentials, but whether the class of origin differentials indicate the pre-mobility differentials is more problematic. The problem is that health was measured at the time of interviewing, when the respondents were 30-69 years of

age. However, the beginning of the intergenerational mobility sequence should be considered to be the point in respondents' lives when they were about to leave their fathers' occupational class and start their own adult career. The health among respondents at that time, in late adolescence/early adulthood, should be the basis for estimating pre-mobility health differentials.

We can estimate the pre-mobility height differential from these data. The special characteristic of height is that it is determined in adolescence. Environments during adulthood do not influence a person's height. Certainly, social conditions during childhood and adolescence, for instance nutrition, may influence bodily growth,[23] and may be one reason for the average lower stature among those from a working-class background. However, after adolescence height becomes a fixed characteristic which is not influenced by later social conditions.

Therefore, the class of origin height differential indicates the pre-mobility height differential and the change in height differential, from class of origin to class of destination, arises only as a consequence of how mobility has redistributed respondents with varying height in the class structure. Accordingly, the slight change in height differential between the worker and higher occupational categories, from 1.6 cm according to class of origin to 1.9 cm according to class of destination (*table 3*), must be produced by intergenerational social mobility. This increase is not statistically significant, but gradient constraint, i.e. a reduced height differential in consequence of mobility, appears not to be present in these data.

As regards the other health indicators, interpretation is more difficult. Since they had left their fathers' homes, the respondents had been exposed to varying degrees of health-beneficial and health-damaging environments, i.e. to social causation. It is likely that the reports of self-perceived health, somatic symptoms, mental symptoms and medical diagnoses were influenced by this social causation during the adult life course of the respondents.

Therefore, using the class of origin differentials to estimate the pre-mobility differentials as regards these four indicators is problematic. We may of course suppose that the measurements made when the respondents were 30-69 years of age reflect pre-mobility health. However, as the measurements probably also reflect experiences of adulthood, the class of origin differentials are uncertain indicators of what the pre-mobility health differentials were in early adulthood. For this reason, determining exactly how the differentials have changed from the pre-mobility to post-mobility situation is problematic and,

because of this, it is difficult to answer whether gradient constraint, i.e. health-related mobility which constrains health differentials, is present as regards these four indicators.

It can be added that, even when trustworthy pre-mobility and post-mobility differentials are available, this does not imply that the role of mobility in changing the differentials is easily found. Differentials can change from a pre-mobility to a post-mobility situation because of social mobility or social causation or both. When we have health indicators which do not change after adolescence (such as height), we can disregard social causation. Normally, however, the health indicators we use are subject to social causation during the mobility sequence we study and only if we have reasonable assumptions about how social causation is involved can the specific effects of social mobility be estimated.

CONCLUSION

Health-related mobility may both widen and reduce social class health differentials. Gradient constraint, i.e. health-related mobility which results in smaller social class differences in health, is likely when the initial health differentials are substantial, mobility is widespread and the association between health and mobility is relatively weak. Changes in health differentials are usually influenced not only by social mobility but also by social causation. Without knowing the role of social causation, the specific effect of health-related mobility is difficult to establish. This empirical study analysed Norwegian data on intergenerational mobility and health. Gradient constraint seems not to be present as regards the height differential. As to the class differentials in the other health variables, the data do not allow for separating the effects of social causation from the effects of health-related mobility and the specific effects of mobility are therefore difficult to determine.

Acknowledgements

This work was funded by the Research Council of Norway and NOVA. Data were collected by Statistics Norway and made available for research by Norwegian Social Science Data Services. Neither of these institutions are responsible for the use and interpretation presented here. Espen Dahl, Institute for Applied Social Science, Oslo and two anonymous reviewers are thanked for their valuable comments.

REFERENCES

1. Hollingshead AB, Ellis BR, Kirby E. Social mobility and mental illness. Am Sociol Rev 1954;19:577-84.

2. Goldberg EM, Morrison SL. Schizophrenia and social class. Br J Psychiatr 1963;109:791-5.

3. Illsley R. Occupational class, selection and the production of inequalities in health. Quart J Soc Affairs 1986;2(2):151-65.

4. Wadsworth MEJ. Serious illness in childhood and its association with later-life achievement. In: Wilkinson RG, editor. Class and health. Research and longitudinal data. London and New York: Tavistock Publications, 1986:50-74.

5. West P. Rethinking the health selection explanation for health inequalities. Soc Sci Med 1991;32:373-84.

6. Dahl E. High mortality in lower salaried Norwegian men: the healthy worker effect? J Epidemiol Commun Hlth 1993;47:153-7.

7. Faresjo T, Svardsudd K, Tibblin G. Social mobility and health in a prospective study of middle-aged men. Scand J Soc Med 1994;22:86-9.

8. Elstad JI. Employment status and women's health – exploring the dynamics. Acta Sociol 1995;38(3):231-49.

9. Miech RA, Caspi A, Mofitt TE, Wright BRE, Silva PA. Low socioeconomic status and mental disorders: a longitudinal study of selection and causation during young adulthood. Am J Sociol 1999;104(4):1096-131.

10. Stern J. Social mobility and the interpretation of social class mortality differentials. J Soc Policy 1983;12:27-49.

11. Townsend P, Davidson N. Inequalities in health: The Black Report. London: Pelican, 1982.

12. Wilkinson, RG. Socio-economic differences in mortality: interpreting the data on their size and trends. In: Wilkinson RG, editor. Class and health. Research and longitudinal data. London and New York: Tavistock Publications, 1986:1-20.

13. Fox JW. Social class, mental illness and social mobility – the social selection drift hypothesis for serious mental illness. J Hlth Soc Behav 1990;31:344-53.

14. Blane D, Davey Smith G, Bartley M. Social selection: what does it contribute to social class differences in health? Sociol Hlth Illness 1993;15(1):1-15.

15. Dahl E. Sosial ulikhet i helse: artefakter eller seleksjon? (Social inequalities in health: artefacts or selection?) [dissertation]. Oslo: FAFO, 1994.

16. Evans RG, Barer ML, Marmor TR, editors. Why are some people healthy and others not? The determinants of health of populations. Berlin/New York: De Gruyter, 1994.

17. Link BG, Phelan J. Social conditions as fundamental causes of disease, J Hlth Soc Behav 1995; (Extra Issue):80-94.

18. Wilkinson RG. Unhealthy societies. The afflictions of inequality. London and New York: Routledge, 1996.

19. Rahkonen O, Arber S, Lahelma E. Health-related social mobility: a comparison of currently employed men and women in Britain and Finland. Scand J Soc Med 1997;25(2):83-92.

20. Van de Mheen H, Stronks K, Mackenbach JP. A lifecourse perspective on socio-economic inequalities in health: the influence of childhood socio-economic conditions and selection processes. Sociol Hlth Illness 1998;20(5),754-77.

21. Bartley M, Plewis I. Does health-selective mobility account for socioeconomic differences in health? Evidence from England and Wales, 1971 to 1991. J Hlth Soc Behav 1997;38:376-86.

22. Blane D, Harding S, Rosato M. Does social mobility affect the size of the socioeconomic mortality differential? Evidence from the Office for National Statistics Longitudinal Study. J R Stat Soc 1999;162 (Part 1):59-70.

23. Blane D, Davey Smith G, Hart C. Some social and physical correlates of intergenerational social mobility: evidence from the West of Scotland Collaborative Study. Sociology 1999;33(1):169-83.

24. Wikholm S, Hildrum A. Helseundersøkelsen 1995. Dokumentasjon og frekvensfordelinger (Health Survey 1995. Documentation and frequencies). Bergen: Norwegian Social Science Data Services, 1997.

25. Album D. Standard for inndeling etter sosioøkonomisk status (Standard classification of socioeconomic status). Oslo-Kongsvinger: Statistics Norway, 1984.

26. Dahl E, Birkelund G. Sysselsetting, klasse og helse 1980-1995 (Employment, class and health 1980-1995). Tidsskrift for Samfunnsforskning 1999;40(1):3-32.

27. Derogatis LR, Cleary PA. Confirmation of the dimensional structure of the SCL-90. A study in construct validation. J Clin Psychol 1977;33(4):981-9.

28. Goldberg DP, Rickels K, Downing R, Hesbacher P. A comparison of two psychiatric screening tests. Br J Psychiatr 1976;129:61-7.

Table 1 Effects of health-related mobility on the size of the health differential: hypothetical examples with two classes, two health statuses, initially 20% health differential and varying mobility probabilities

	Initial situation	Mobility		Result after mobility
		Stable	Movers	
Example A				
Upper class	400 g	280 g	120 g	520 g
	100 b	60 b	40 b	180 b
Bad health (%)	20	18	25	26
Lower class	600 g	360 g	240 g	480 g
	400 b	280 b	120 b	320 b
Bad health (%)	40	44	33	40
Example B				
Upper class	400 g	320 g	80 g	560 g
	100 b	60 b	40 b	140 b
Bad health (%)	20	16	33	20
Lower class	600 g	360 g	240 g	440 g
	400 b	320 b	80 b	360 b
Bad health (%)	40	47	25	45

Example A. Mobility probabilities: down from upper class 0.4 if bad health and 0.3 if good health and up from lower class 0.3 if bad health and 0.4 if good health.

Example B. Mobility probabilities: down from upper class 0.4 if bad health and 0.2 if good health, and up from lower class 0.2 if bad health and 0.4 if good health.

g, good health; b, bad health.

Table 2 *Class of origin and class of destination, 1,853 men aged 30-69 years, Health Survey 1995, Norway (row percentages with number of respondents in parentheses)*

	Occupational class of destination			
Class of origin	Workers	Intermediate	Higher	Total %
Workers	43.3 (332)	14.0 (107)	42.8 (328)	100.1 (767)
Intermediate	30.8 (226)	28.1 (206)	41.1 (301)	100.0 (733)
Higher occupations	15.0 (53)	14.4 (51)	70.5 (249)	99.9 (353)
Total %	33.0 (611)	19.6 (364)	47.4 (878)	100.0 (1,853)

Table 3 *Mean values on height, self-perceived health (proportion below good), somatic symptoms, mental symptoms and medical diagnoses, age-adjusted, different stages of the mobility process, men aged 30-69 years, Health Survey 1995, Norway*

	Height (cm)	Self-Perceived Health	Somatic symptoms	Mental symptoms	Medical diagnoses
Class of origin					
Father worker	179.1	0.200	0.560	0.628	1.602
Father intermediate	178.9	0.175	0.570	0.534	1.458
Father higher occupation	180.7	0.105	0.378	0.529	1.432
Class of destination					
Own class worker	178.3	0.234	0.597	0.704	1.650
Own class intermediate	179.2	0.211	0.720	0.807	1.623
Own class high occupation	180.2	0.123	0.410	0.407	1.396
Differential between Workers/Higher occupation					
Class of origin	–1.6	+0.095	+0.182	+0.099	+0.170
95% CI lower	–2.4	+0.052	+0.057	–0.126	–0.038
95% CI upper	–0.8	+0.138	+0.308	+0.325	+0.378
Class of destination	–1.9	+0.111	+0.187	+0.297	+0.254
95% CI lower	–2.6	+0.082	+0.084	+0.115	+0.148
95% CI upper	–1.3	+0.161	+0.294	+0.497	+0.494
Change	+0.3	+0.016	+0.005	+0.198	+0.084
95% CI lower	–0.7	–0.043	–0.159	–0.144	–0.226
95% CI upper	+1.3	+0.075	+0.169	+0.540	+0.394
Stable members/Movers Worker/Higher occupation					
Left working class	179.7	0.159	0.533	0.517	1.527
Stable work class	178.4	0.259	0.600	0.785	1.735
Enter work class	178.1	0.206	0.593	0.609	1.552
Left higher occupation	179.6	0.144	0.407	0.885	1.515
Stable higher occupation	181.2	0.091	0.371	0.392	1.405
Enter higher occupation	179.8	0.137	0.433	0.408	1.404