Social Determinants of Health - Contributions from European Health and Medical Sociology

Determinantes sociales de la salud - Contribuciones de la sociología médica y de la salud europea

Johannes Siegrist
Department of Medical Sociology. University of Duesseldorf
siegrist@uni-duesseldorf.de

Recibido: 25.10.2010
Aprobado definitivamente: 15.11.2010

RESUMEN
Uno de los principales temas de las investigaciones recientes de la sociología médica o de la salud tiene que ver con los determinantes de la salud, o más específicamente con las inequidades sociales en salud. Sociólogos de la medicina o epidemiólogos sociales europeos han realizado sustanciales aportes al respecto. En este artículo realizo una breve revisión del actual estado de la cuestión sobre el gradiente social en salud, teniendo como referencia cinco aproximaciones complementarias: selección social, provisión de servicios de salud, penurias en los primeros años de vida, estilos de vida poco saludables, situaciones estresantes/recursos materiales y psicosociales en la vida adulta. Esta última aproximación se explicará de manera más detallada en relación con los efectos adversos rara la salud de las condiciones de trabajo y empleo. Se presentan, además, nuevas evidencias sobre la influencia que tienen el deterioro de las condiciones de trabajo sobre la salud física y mental, haciendo especial referencia a dos modelos sociológicos que dan cuenta de las situaciones de estrés laboral: demanda-control y desajustes entre el esfuerzo y recompensas obtenidas. Por otra parte, se plantea la discusión de la relación entre condiciones de trabajo y salud en la vida adulta en el contexto de dos hipótesis: la mediación y la modificación de hipótesis. Como conclusión, se plantea que en los últimos años se ha logrado un progreso importante de la subespecialidad de la sociología médica o de la salud, y este nuevo conocimiento sociológico ya está siendo aplicado en diferentes contextos de la salud y de la política social.

PALABRAS CLAVE: Reformas sanitarias, gestión sanitaria, profesión médica, neoliberalismo, gubernamentalidad, desprofesionalización, organizaciones sanitarias

ABSTRACT
A major domain of recent research in health or medical sociology is concerned with social determinants of health, and more specifically with the explanation of social inequalities in health. Substantial input to this research was provided by European medical sociologists and social epidemiologists. In this article, I give a short review of the current state of art in explaining the social gradient of health, pointing to five complementary approaches: social selection, health care provision, early life deprivation, unhealthy lifestyles, and material and psychosocial stressors/resources in adult life. This latter explanation is discussed in more detail with regard to health-adverse working and employment conditions. New scientific evidence on the influence of poor quality of work on physical and mental health is presented, with particular reference to two sociological models of stressful work, demand-
control and effort-reward imbalance. Moreover, their contribution towards explaining unequal adult health is discussed in the frame of two hypotheses, the mediation and the effect modification hypothesis. In conclusion, substantial sociological input and progress of the sub-specialty of health or medical sociology was achieved in recent past, and this new knowledge is already being applied in various contexts of health and social policy.

**KEYWORDS:** Health inequalities, medical sociology, workstress, effort-reward imbalance, health policy.

**SUMARIO**

INTRODUCTION

Since its beginning in the nineteenth century, sociology has often been defined as a science whose aims are not restricted to the description, measurement and explanation of specific aspects of reality – the structures and processes of functioning of societies. Rather, again and again, sociologists have been committed to the additional tasks of criticizing societal life on the basis of their scientific findings and of promoting societal change. These additional tasks seem well justified given the fact that avoidable injustice and inequality prevail, to a larger or lesser extent, in every human society. This combination of scientific analysis with critical judgment and normative action is not restricted to general sociology, but applies equally to its many sub-specialties including health or medical sociology.

Health or medical sociology is the sub-specialty of general sociology that deals with two broad health-related areas. On the one hand, it is concerned with the sociological analysis of health care institutions, their professional groups and clients as well as the larger socioeconomic and cultural frames of health care delivery. This field has been labelled ‘sociology of medicine’. On the other hand, social determinants of health are the main focus of analysis where societal effects on the onset and course of diseases and on the promotion of health are explored (‘sociology in medicine’, ‘sociology of health’). Although there are good reasons to criticize this taxonomy it proofs to be useful in supporting the development of cumulative knowledge within the broader field of sociological inquiry (Cockerham 2009). This holds particularly true for the second task mentioned, the study of social determinants of health which are the subject of this contribution.

Research on social determinants of health has witnessed a surprisingly dynamic development over the past three or four decades, originating largely from distinct research centres in Europe, the USA and Canada (Berkman & Kawachi 2000, Marmot & Wilkinson 2006, Siegrist & Marmot 2006). Meanwhile, a growing body of theory-based internationally comparable knowledge has been produced with a particular focus on social inequalities in health. In line with the specific features of sociological inquiry, the results of this research have stimulated a critical appraisal of avoidable health inequalities, and they are now instrumental in supporting policies of societal change in the field of health (WHO 2008, The Marmot Review 2010). It is true that these achievements are mainly attributable to the work of social epidemiologists, but it is equally true that their main theoretical input was developed by health or medical sociologists (see below). In fact, scientific collaboration of epidemiologists and sociologists has resulted in a particularly successful type of inter-disciplinary cooperation.

In the following sections I first give a short review of the current state of knowledge about main explanations of social inequalities in health, with a special focus on the European dimension. This short review is followed by a more detailed discussion of one particular aspect of the broader field of social determinants of health, the role of work and employment in producing unequal health. The concluding remarks deal with the contribution of sociological theory to this domain of research and with some policy implications of the current state of knowledge in this field.

STEPS TOWARDS EXPLAINING SOCIAL INEQUALITIES IN HEALTH

Despite considerable progress in medical science, constant increase in health care spending and extension of health care facilities as part of modern welfare regimes spread across Europe substantial health inequalities still prevail in all European countries from which respective data are available (Mackenbach & Bakker 2002, Siegrist & Marmot 2006). With regard to mortality, mean difference in life expectancy between those at the top and those at the bottom of a society’s social structure, as defined by education, income, or employment status, are anywhere from four to ten years. Several investigations observed a widening of social inequalities in health during the final quarter of the last century (Drever & Whitehead 1997). Importantly, the problem of inequality in health is not confined to the poorest members of society, but there is a social gradient of mortality and morbidity across the whole of a society. With each step one moves up on the social ladder, the better one’s health (Marmot 2004).
A social gradient has been documented for premature mortality and for the incidence and prevalence of all the major causes of death, including coronary heart disease, stroke, certain cancers (e.g. lung cancer), type 2 diabetes, obstructive pulmonary disease, liver cirrhosis, sexually transmitted diseases, depression, and accidents including homicide and suicide (van Rossum et al. 2000). Moreover the prevalence of handicap and impairment is socially graded, and differences in healthy life expectancy are much steeper than differences in expected years of life.

Research has shown that the social gradient of health varies across the life course where it is steepest at two stages, early childhood and midlife, whereas less inequality is observed in adolescence and in older age (Kuh & Ben Shlomo 2004). Moreover, variations in health occur according to gender, with a more explicit gradient in men than in women, and according to country or region. Most obviously, a substantial gap in life expectancy exists between Eastern and Central European countries on one side, and Western Europe on the other side (Marmot 2004). It should also be noticed that the social gradient differs according to the measure of inequality applied to the study. In general, most consistent associations are observed with regard to education as an indicator of social status (Mackenbach & Bakker 2002), but among middle-aged men, occupational position reflects the differential susceptibility to morbidity and mortality better than alternative indicators (Siegrist & Marmot 2006).

What are the main explanations of this social gradient in health? Research so far has provided five interrelated answers. First, it has been claimed that natural or social selection account for health inequalities. This explanation posits that health ‘selects’ people into different social strata, thus reversing the direction of causation. Recent results from birth cohort studies show that the contribution of poor health during childhood on social mobility later in life is real, but small and that the causal direction is likely to be from social environment to illness, not the other way (Kuh & Ben Shlomo 2004). A second explanation points to different access to, and quality of, medical care according to socioeconomic position. The relevance of this explanation clearly varies according to the characteristics of a health care delivery system. For instance, it is more powerful in a country with low availability of health insurance, such as the United States of America, than in a country with full health care coverage of the population, as in Scandinavian welfare states and other European countries. Third, social deprivation and adversity at the beginning of life, during pregnancy and during early childhood, were shown to have negative short- and long-term effects on health (Power & Kuh 2006). These effects are due to a variety of environmental and parental influences, including material, behavioural and psychosocial factors. A fourth, very powerful explanation concerns health-adverse behaviours. To a large extent, socioeconomic and socio-cultural influences shape health-related behaviours during primary socialization and peer group exposure in adolescence, in particular poor diet, smoking, alcohol consumption, lack of physical exercise and overweight. A social gradient of this unhealthy lifestyle has been documented in many European countries (Marmot & Wilkinson 2006). When exploring the reasons behind these graded health lifestyles one may refer to Pierre Bourdieu’s (1984) concept of ‘habitus’ that points to the correspondence between socio-structural opportunities or constraints and practices of everyday life in different domains. Through the processes of socialization, these practices come to be embodied as schemes of perceiving, feeling and acting, thus giving rise to characteristic behavioural dispositions, such as health-related patterns of behaviour that are stratified according to people’s social standing (Singh-Manoux & Marmot 2005).

Finally, a fifth approach towards explaining the social gradient of health claims that material and psychosocial constraints and resources related to core social roles in adult life have a far-reaching impact on health. Work and employment, marriage and parenthood, and participation in civic life define such core roles, and it is the availability and quality of these roles that triggers health via material and psychosocial pathways. The materialist (or neo-materialist) explanation posits that material resources related to income have a direct impact on health. As income determines a wide range of life circumstances one can assume that conditions such as poor housing, lack of recreation, or cheap and unhealthy nutrition all contribute to poor health. Yet, the
materialist explanation fails to account for health variations in less disadvantaged parts of modern societies where a broad spectrum of psychosocial influences on health interacts with material conditions. Here, factors such as social participation and integration (often referred to as 'social capital'), quality of social relationships and of working conditions, and availability of skills and competencies in coping with the challenges of life seem to matter more than material factors (Siegrist & Marmot 2006, Wilkinson & Pickett 2009).

In the following section, this latter argument is discussed in more detail with regard to the work role, a core social role in adult life. Yet, we should stress the fact that the explanations of social inequalities in health mentioned are interrelated, pointing to a complex web of causation with cumulative and moderating effects acting at different periods of time over the life course. Moreover, more recent sociological research has demonstrated that within each one of these explanations one needs to distinguish the contextual influences triggered by social stratification from the influences that have direct impact on individual behaviour, attitudes, and emotions. Respective evidence comes from applying multi-level analysis to the study of social determinants of health (Kawachi & Berkman 2003).

In summary, while substantial progress in sociological research on social determinants of health was achieved in recent past, several challenges remain to be overcome by future scientific inquiry both at the conceptual and methodological level. One strategy of dealing with these challenges concerns the meso-social level of sociological analysis. It proposes an in-depth study of core psychosocial environments to which people are exposed in their everyday life. As an example, I illustrate this strategy with regard to the psychosocial work environment.

QUALITY OF WORK AND EMPLOYMENT AND UNEQUAL HEALTH

Occupational research has long been concerned with the material dimensions of health-adverse work, such as heat, noise, physical or chemical hazards at the workplace. While these hazards are still highly relevant in specific occupational groups, large proportions of the work force in modern economies are exposed to mental and emotional demands and threats at work, rather than material demands and hazards. As a result, psychological and social stressors are becoming more frequent, and their contribution to health and well being at work is likely to outweigh the contribution of more traditional occupational stressors. In this context the term ‘psychosocial work environment’ was introduced and defined as the socio-structural range of opportunities that is available to a working person to meet his or her needs of well being, productivity and positive self-experience, in particular self efficacy and self esteem (Siegrist & Marmot 2006). Thus, more recent research was devoted to advance explanations of social inequalities in health in modern societies by focusing on the impact of an adverse psychosocial work environment.

To identify a psychosocial work environment and to demonstrate its effects on health theoretical models are needed. A theoretical model is best understood as a heuristic device that selectively reduces the complex reality to meaningful components. These components are delineated at a level of generalization that allows for their application to a wide range of different phenomena, and in this case a wide range of different occupations. A theory is commonly defined as a set of interrelated statements that explain the statistically documented associations, in this case the associations between working conditions and health, by elucidating underlying pathways.

A variety of theoretical models of a health-adverse psychosocial work environment were developed (for review e.g. Cartwright & Cooper 2009), but few only were widely tested in the frame of prospective epidemiological investigations with clinical disease outcomes. Among these, two models received particular attention in international research during the past two decades, the demand-control model and the effort-reward imbalance model.

The demand-control model (Karasek & Theorell 1990) posits that stressful experience at work results from a distinct job task profile defined by two dimensions, the psychological demands put on the working person and the degree of control available to the person to
perform the required task. Jobs defined by high demands and low control are stressful because they limit.

The individual’s autonomy and sense of control while generating continued pressure (‘high job strain’). A third dimension, social support at work, was added to the original formulation where highest strain is expected to occur in jobs that are characterized by high demand, low control and low support at work or social isolating (‘iso-strain jobs’; Johnson & Hall 1988).

The effort-reward imbalance model is concerned with stressful features of the work contract (Siegrist 1996) (see Figure 1). This model builds on the notion of social reciprocity, a fundamental principle rooted in an ‘evolutionary old’ grammar of interpersonal exchange. Social reciprocity lies at the core of the employment (or work) contract which defines distinct obligations or tasks to be performed in exchange with adequate rewards. These rewards concern money, esteem, and career opportunities including job security. Contractual reciprocity operates through norms of return expectancy, where efforts spent by employees are reciprocated by equitable rewards from employer. The effort-reward imbalance model claims that failed reciprocity in terms of high cost and low gain elicits strong negative emotions with special propensity to sustained autonomic and neuroendocrine activation and their adverse long-term consequences for health. Contractual non-reciprocity is expected to occur if one or several of the following conditions are given: ‘dependency’, ‘strategic choice’, and ‘overcommitment’. Dependency reflects the structural constraints where no alternative choice in the labour market is available. Strategic choice defines situations of heavy competition where high efforts without adequate compensation are invested in order to improve future chances of career promotion. Overcommitment points to psychological reasons of this mismatch as people characterized by a motivational pattern of excessive striving at work may exceed the level of expected performance because of their underlying need for approval and esteem.

**Figure 1**
The model of effort-reward imbalance (J. Siegrist 1996)
In a variety of epidemiological studies, each one of these models was shown to predict elevated risks of stress-related physical and mental disorders. Most investigations were designed to predict coronary heart disease or depression as disease outcomes, given their public health importance. At least 15 prospective observational studies tested the demand-control and/or the effort-reward imbalance model with regard to cardiovascular diseases. In a majority of cases, significantly elevated relative risks were observed with an overall increased probability ranging between 50 and 80 per cent (Kivimäki et al. 2006). A somewhat higher and more consistent association was observed with regard to depression where people exposed to high demand and low control or effort-reward imbalance at work exhibited a twofold elevated risk of affective disorder (Stansfeld & Candy 2006, Kivimäki et al. 2007).

Additional health outcomes were studied, such as poor self-rated health, limited physical and mental functioning, alcohol dependence, and risk of type 2 diabetes (Chandola et al. 2008, Siegrist 2009). It should be noted that results are drawn from a variety of different occupations in different modern and rapidly developing societies, thus representing rich empirical support. Moreover, naturalistic and experimental studies supplement epidemiological evidence as they monitor the psychobiological responses to stressful work, such as cardiovascular reactivity, hormonal secretion, immune function or inflammatory response (Steptoe 2006).

Having demonstrated that an adverse psychosocial work environment in terms of these two theoretical models affects the physical and mental health of working people we now ask what this research contributes to the explanation of social inequalities in adult health. Three answers are of interest here. First, in general, a higher prevalence of stressful work is observed among people with low educational training and among those holding low occupational positions. This is particularly often the case with regard to high demand-low control jobs that are typical of many jobs in mass industry (e.g. piece work and machine paced assembly line) and in the low paid service sector (e.g. waiters, operators in call centres).

Associations of effort-reward imbalance at work with socioeconomic position are less consistent because of a high frequency of demanding jobs in employees with high socioeconomic status. However, concerning all three dimensions of occupational rewards, a clear social gradient is given.

If stressful work is more prevalent in lower socio-economic groups and if both, work stress and low socio-economic position, are associated with reduced health, it is tempting to conclude that work stress mediates the association of socio-economic status with health. Therefore, the second answer points to available evidence in favour of this mediation hypothesis. Several studies tested this hypothesis using multivariate logistic regression analysis with consecutive model estimates. Some of their findings indicate that the statistical association of socioeconomic status with health, as expressed in the size of odds ratios, is attenuated if variables measuring an adverse psychosocial work environment are included into the multivariate model. For instance, in the well known Whitehall II study of British civil servants low control in the workplace accounted for about half of the social gradient of coronary heart disease as the odds ratio of coronary disease in the low status employment group was reduced from about 1.4 to about 1.2 after respective adjustment (Marmot et al. 1997). Yet, the mediation hypothesis was not confirmed in some other studies, thus requiring a complementary third answer to the question raised above. This answer is provided by the effect modification hypothesis. It posits that susceptibility to an exposure is higher among lower-status compared to higher-status people and, therefore, that among the former, the effect size produced by the exposure is higher. In our case, being exposed to a high-demand job or a job defined by high cost and low gain is more stressful for people with low social standing than for those who are better off, and therefore, the odds ratios of stress-related disorders are generally higher among lower status people. This may be due to their higher overall burden of stressful living and working conditions or their reduced psychosocial resources of coping with adversity. Susceptibility may also be increased as a result of a higher prevalence of unhealthy lifestyles or more advanced sub-clinical disease development (Siegrist & Marmot 2006). For example, in a cross-sectional epidemiological
study the probability of experiencing depressive symptoms was eight times higher in a group of workers with low occupational status and high work stress in terms of effort-reward imbalance compared to a group of employees with high social standing and low or no stress at work (Wege et al. 2008).

In summary, an adverse psychosocial work environment contributes to the burden of stress-related diseases in adult life, and the theoretical models assessing this adverse psychosocial environment offer new explanations of the social gradient of health in terms of the mediation hypothesis and the effect modification hypothesis. Clearly, this meso-social level of analysis has to be complemented by macro-level analyses that address the larger labour market, economic and technological challenges given in the era of globalization (Schnall et al. 2009).

CONCLUDING REMARKS

In this contribution I set out to demonstrate that the sub-specialty of health or medical sociology made substantial scientific progress in recent past, in particular in the domain of studying social determinants of health. This progress was conditional on collaborative efforts jointly with social epidemiologists, but relevant theoretical input relied on contributions from the discipline of sociology. This became obvious when current explanations of the social gradient of health were discussed. In fact, several theoretical concepts from classical and modern sociology were successfully applied to this field of research, notably the concepts of social stratification (Weber 1978, Antonovsky 1967, Hollingshead & Redlich 1958), social roles (Merton 1967, Kasl & Cobb 1970) and social integration (Durkheim 1951; Berkman & Syme 1979). More recent examples are the concepts of habitus (Bourdieu 1984), social capital (Coleman 1990, Putnam 2000), autonomy and control (Kohn & Schooler 1983, Karasek 1979), or social reciprocity and social reward (Gouldner 1960, Siegrist 1996). It should also be noticed that the fundamental discovery of Emile Durkheim (1950) that society is not simply the sum of individuals but consists of genuine collective features was instrumental in promoting a multitude of recent multi-level analyses in the field of social determinants of health (Kawachi & Berkman 2003). These promising developments need to be carried on and extended.

In the Introduction, it was also mentioned that sociology as a scientific discipline is concerned with the implications of new knowledge for societal change and progress. In this regard, health or medical sociology, and again most obviously by its collaboration with social epidemiology, made significant contributions to health and social policy. For instance, a series of evidence-based recommendations was elaborated on how to reduce the social gradient of morbidity and mortality in fields such as income, tax and benefits, education, housing, work and employment, nutrition, different age groups, and health care delivery (Marmot 2004). Health policies based on scientific knowledge about social determinants of health are currently applied in several member states of the World Health Organisation (WHO 2008). In addition to these macro-political efforts, more specific targets of reducing unequal health are tackled at the meso-social level, specifically in the frame of worksite health promotion initiatives (The Marmot Review 2010). Reducing psychosocial stress at work by implementing theory-based measures of organisational and personnel development is an important target of reducing unequal health in midlife and early old life and of extending the employment capability of a growing older workforce (Schnall et al. 2009).

In conclusion, health or medical sociology continues to be a vital field of scientific inquiry with far-reaching policy implications. Significant contributions to its growth were achieved in distinct European countries, and some these European countries are also pioneering in the application of new sociological knowledge to the development and implementation of health and social policy programmes.
REFERENCES


DURKHEIM E (1951). The Suicide. New York: The Free Press (original in French, 1897)


