

STRATEGIES FOR HEALTH

AN ANTHOLOGY

STRATEGIES FOR HEALTH

AN ANTHOLOGY

EDITED BY PER NILSEN



Linköping University

Linköpings universitet
Institutionen för Hälsa och Samhälle (IHS)
Department of Health and Society
S-58183 Linköping
Sweden

Published 2007

Copyright by Hanna Arneson, Kerstin Ekberg, Per Nilsen, Lennart Nordenfelt, Elsy Söderberg

Cover design by Aaron A. Sikkink
Text editing by Diana Stark-Ekman and Per Nilsen
Printed by LiU-tryck, Linköping, Sweden, 2007

All rights reserved. Except for the quotation of short passages for the purposes of criticism and review, no part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form, or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publisher.

ISBN: 978-91-85715-36-7

ISSN: 1652-1994

IHS Rapport 2007:1

CONTENTS

PREFACE	1
UNDERSTANDING THE CONCEPT OF HEALTH <i>Lennart Nordenfelt</i>	4
FROM “ACTS OF GOD” TO SAFETY PROMOTION – A BRIEF HISTORY OF INJURY RESEARCH <i>Per Nilsen</i>	16
THE THEORY OF COMMUNITY-BASED HEALTH AND SAFETY PROGRAMMES – A CRITICAL EXAMINATION <i>Per Nilsen</i>	28
WORKPLACE HEALTH – INFLUENCES AND INTERVENTIONS <i>Kerstin Ekberg</i>	46
EMPOWERMENT AND HEALTH ENHANCEMENT IN WORKING LIFE – FRAMING THE CONCEPT, REVIEWING THE EVIDENCE <i>Hanna Arneson</i>	59
SOCIETY-BASED STRATEGIES AND MANAGEMENT ROUTINES AIMED AT PROMOTING RETURN TO WORK <i>Elsy Söderberg</i>	71

PREFACE

The ever-growing complexity of modern society has required that researchers expand their collaboration across scientific disciplines. Interdisciplinary research has the potential to yield insights and knowledge that cannot be achieved by research within an individual discipline. Although each discipline has its own concepts, theories, models, and methodologies that it uses to answer questions, the convergence of perspectives from multiple disciplines may stimulate new, creative approaches to solving challenging research problems.

The need for interdisciplinary research was a motivating factor in the establishment of Institutionen för Hälsa och Samhälle (IHS) (the Department of Health and Society) in 2002 at Linköping University, Sweden. IHS encompasses medical, philosophical, and technical faculties, and is composed of six divisions: Socialmedicin och folkhälsovetenskap (Social Medicine and Public Health Science), Rikscentrum för arbetslivsinriktad rehabilitering (National Centre for Work and Rehabilitation), Sjukgymnastik (Physiotherapy), Allmänmedicin, Centrum för utvärdering av medicinsk teknologi, and Tema hälsa och samhälle (Tema Health and Society).

Encouraged by the multidisciplinary research approach of IHS, scientists at the six divisions participate in three strategic research arenas. These arenas offer researchers from a variety of backgrounds a forum to expand their vocabulary to understand each other and to develop and share concepts, theories, models and methodologies, thus facilitating interdepartmental research. One of the three research arenas is Strategies for Health, which is comprised of researchers from the five divisions of the medical and philosophical faculties of IHS.

Strategies for Health research recognises that health is a broad concept, encompassing physical, mental, and spiritual dimensions. Health is fundamentally interlinked with the physical, social, cultural, and economical environments surrounding people. This means that health can be promoted on different levels, from individual to organisational and societal levels, and in arenas where people live, work, and play, e.g. workplaces, schools, neighbourhoods, and communities.

Strategies for Health research addresses the interaction among society's many welfare actors and the strategies and policies they implement to influence the health of individuals and populations. Intervention research is an integral element of Strategies for Health, with the aim of producing knowledge about problem analysis, design, implementation, and evaluation of interventions in order to achieve effective and sustainable solutions. Another aspect of Strate-

gies for Health research investigates socio-cultural beliefs and ideas individuals and groups have concerning their abilities and opportunities to attain health, and the strategies they use in order to improve and/or maintain their health.

This anthology is comprised of six papers, introducing readers to a variety of topics relevant to the Strategies for Health arena. The papers are intended to reflect the breadth of research conducted in Strategies for Health. The aim is to contribute to a greater understanding of the concept of health and provide insights into some strategies for improvement of health and safety.

The opening paper by Lennart Nordenfelt, “Understanding the concept of health,” sets the scene by stating that health is not simply the absence of disease, but a multidimensional concept. Nordenfelt explores a multitude of characterisations of health, comparing the medical (biostatistical) and holistic conceptions of health and disease. He argues that a holistic orientation to health is most appropriate since health in medical practice and public health contexts is considered something beyond the absence of or prevention of disease. Thus, health is best interpreted along holistic lines of reasoning.

A holistic definition of health implies that there is a broad range of strategies to promote health. The field of safety promotion has evolved from behavioural strategies predominantly aiming at changing the individual’s behaviour, to strategies that place more emphasis on modification of physical and social environments, often integrating different approaches in broad, multifaceted community-oriented programmes to promote safety at the population level. The paper “From ‘acts of God’ to safety promotion – A brief history of injury research” by Per Nilsen chronicles injury research’s evolution, from its “pre-scientific” era, when injuries were considered unpreventable random events, i.e. “acts of God,” to the transformation of injury epidemiology into a science over the latter half of the twentieth century.

The community-based approach to health and safety promotion represents a shift in focus from individual accountability and “victim blaming” to explanations that also encompass social and environmental influences. However, despite wide application of community-based programmes during the last 30 years, there is a paucity of evaluations from which to obtain evidence regarding the effectiveness of these programmes. In “The theory of community-based health and safety programmes – A critical examination,” Nilsen scrutinises the theory of the community-based approach to examine whether there are shortcomings in the theoretical underpinnings of this strategy that could explain the lack of convincing evidence of effectiveness.

Communities are complex “open systems,” often making it difficult to reach many subgroups. In contrast, the workplace constitutes a more clearly delineated environment, with ease of access to a large proportion of the adult population and established channels of communication. However, the workplace brings together many groups and individuals who may have different priorities

with regard to health and work, potentially yielding a conflict between the pursuit of productivity and improvement of health. In “Workplace health – Influences and interventions,” Kerstin Ekberg reviews key concepts, theories, and models to explain the relationship between health and work. She also discusses research findings concerning strategies to improve workplace health.

The workplace as an important determinant of health is further explored by Hanna Arneson in “Empowerment and health enhancement in working life – Framing the concept, reviewing the evidence.” Arneson discusses the concept of empowerment and how this has been applied within the context of working life. Empowerment is a concept that is shared across many disciplines, including community development, psychology, education, management, and studies of social movements and organisations, each contributing its own perspective to the concept. In health promotion, empowerment is a foundational concept as it constitutes a positive multi-factorial approach to health, typically being understood as a process through which people can gain greater control over decisions and actions affecting their health. Arneson reviews empirical research findings pertaining to empowerment in working life and its association with health, concluding the paper by outlining a number of key issues that require further research.

Despite the potential of the workplace as a setting for improving health and safety, many Western countries have experienced high and/or increasing levels of sickness absence and rising sickness insurance system costs. This development has prompted societal efforts to restore the health and functional ability of people on sick leave. However, promoting return to work in ill workers is a complex process that involves actions by and interaction between numerous actors, including employers, health and medical services, and social services. In her article, “Society-based strategies and management routines aimed at promoting return to work,” Elsy Söderberg looks at the strategies used by different welfare actors and pinpoints crucial difficulties associated with these efforts. Söderberg concludes her paper, and the entire anthology, by delineating important areas for further research.

Per Nilsen with Lennart Nordenfelt, Kerstin Ekberg, Hanna Arneson, and Elsy Söderberg

UNDERSTANDING THE CONCEPT OF HEALTH

LENNART NORDENFELT

On the value of health

Many people, in particular in modern times, have regarded health as one of the most precious values in life. Health, as well as longevity, should, they think, be protected and enhanced as much as possible. Thus, the art and science of medicine has received a crucial place in the modern, both Western and Eastern, society. Doctors and other health workers are important people. They are highly regarded and they are well paid in most countries. In certain circles they have replaced the priests or even the gods of old times, with the power to prolong life.

Along with the growing stature of health care providers in modern society, we can also see the rapid development of health promotion, a movement which is partly but not wholly connected to the development of medicine. Health promoters of various kinds play roles as advisors and supporters to many modern people. Commercial industry has followed in these steps. Huge amounts of goods that purport to be beneficial for one's health have been marketed and successfully put up for sale.

This marketing of good health has not always been a prevalent social value. During medieval times in Western Europe one's life on earth was not the important life. This life was only a preparation for the eternal life to follow, together with God. Thus, health in this life did not have the utmost value. It was much more important to successfully prepare oneself for the eternal life and thus live in accordance with the duties indicated in the holy literature, in particular the Bible.

Moreover, most philosophers in Western cultures have preached other virtues than the healthy life. The great Plato from the fourth century BC, for instance, said that we should not concentrate our interest and ambition on our own health or on questions on health and disease. When people concentrate on their own health and want to consult a doctor at all times this is a sign of unsound conditions in the state. Neither should doctors be given power over people. People should never leave the responsibility for their lives in the hands of other people.

Today, however, we find in most Western countries a great attention to health matters. In my own country, Sweden, several investigations have been made about this phenomenon. The best known studies have been performed by the Swedish professor of religious philosophy, Anders Jeffner (see Kallenberg *et al.*, 1997). In his studies Jeffner asked a representative sample of Swedes to identify the attributes that they valued most in their lives. A vast majority of these people put health on top of their lists, followed by attributes such as wealth, participation in cultural activities, high social status, good family relations and world peace!

Is there a good explanation of this phenomenon, where health is considered the best attribute one can have in one's life? Indeed, I think there is such an explanation. Sweden is probably the most secular country in the world, even in comparison with the post-communist states, where atheism was officially preached. It is rare that Swedes expect a life after death. Thus, practically all their attention is focused on the problems of this life on earth and on having the best conditions to live this life. Health is, not unexpectedly, believed to be such a condition. Moreover, attainment of good health is now a possibility for many, where, even fifty years ago, such an outcome was well beyond the imagination of many. One must remember that it is only during the last century, because of the development of medicine and health promotion, that it has become possible to make radical improvements in the health status of people. Hygienic conditions in the rich countries are now such that one can usually guarantee a reasonably healthy life to their inhabitants. It has also become possible to cure or prevent some of the most deadly diseases, such as smallpox and tuberculosis. Thus, it is only recently that it has become possible to really hope for a radical improvement of the health states of people in the world.

The purpose of this paper is, however, not to discuss the sociology of today's health interest. I wish instead to contribute to the understanding of the nature of health, which is the area where I have done most of my research. I shall do this mainly by comparing two kinds of philosophies of health which are dominant in the modern discussion. I shall argue in favour of one of these, viz. a holistic understanding of health. It will be evident, I think, that this understanding is the more adequate one, especially in the light of what I have said about people's appreciation of health. Let me however first provide a more general sketch of the philosophy of health.

Some historical theories of health

The varieties of health

Health is a notion primarily applicable to the entirety of a human's well-being. But there are more specific derivative notions of health. Ever since antiquity, reinforced by the Cartesian distinction between body and mind, it has been natural to separate somatic health from mental health. The interpretations of mental health have varied over time. The ancient notion of mental health was closely connected to morality, whereby the mentally healthy person was a person who lived a virtuous life, but this idea has lost most, though not all, of its significance today. The idea of spiritual health is also current in the health science although it is not systematically recognised. Bernhard Häring is a leading spokesman for a notion of health including a spiritual dimension, stating: "A comprehensive understanding of human health includes the greatest possible harmony of all of man's forces and energies, the greatest possible spiritualisation of man's bodily aspect and the finest embodiment of the spiritual" (1987, page 154).

The various categories of health have connections to each other. Sometimes bodily health has been given priority in the sense that it has been viewed as a prerequisite for mental health. Galen (ca. 129-216/7) in some of his writings attempted to explain mental properties of the person in terms of specific mixtures of the bodily parts (Galen, 1997). Consider also the ancient proverb: *mens sana in corpore sano* (a healthy mind in a healthy body). In the modern discussion about mental illness, one position, favoured in particular by medical doctors, is that all mental illness has a somatic background, i.e. that all mental illnesses – if they exist at all – are basically somatic diseases. The customary view, however, also in Western medicine, is that a person can at the same time be somatically healthy and mentally ill, or vice versa.

Health as balance

An extremely powerful idea in the history of medicine is the one that health is constituted by bodily and mental balance. The healthy person is a person in balance, normally meaning that different parts and different functions of the human body and mind interlock harmoniously and keep each other in check. The Hippocratic and Galenic schools (Hippocrates 460–380 BD and Galen 129-210 AD) were the first Western schools to develop this idea in a sophisticated way. They stated that a healthy body is one where the primary properties (wet, dry, cold, hot) of the body balance each other. In the medieval schools, following Galen, this idea was popularised and formulated in terms of a balance between

the four bodily humours: blood, phlegm, yellow bile, and black bile (Cp. humoral pathology).

The idea of balance remains strong in several non-Western medical traditions. The Yajurveda tradition in India, for instance, declares that there are three humours acting in the body: breath (*vata*), bile (*pitta*), and phlegm (*kapha*). The proportions of the three humours vary from person to person, and their actions vary according to season, environment, the life-style of the individual, and his or her diet. In good health the humours are in equilibrium. Disease is the result of their imbalance (Singhal & Patterson, 1993).

Balance is a powerful idea also in modern Western thought, in particular within physiology. The idea is then often to be recognised under the label of *homeostasis* (the Greek word for balance). Walter Cannon's (1871-1945) classical work on homeostasis (1930) describes in detail how the various physiological functions of the body control each other and interact in feedback loops in order to prevent major disturbances.

The idea of balance or *equilibrium* (the Latin word for balance) has a rather different interpretation in the writings of Ingmar Pörn. Here balance is a concept pertaining to the relationship between a person's abilities and his or her goals. The healthy person, according to Pörn, is the person who can realise his or her goals and thus retain a balance between abilities and goals (see Health as ability, below).

Health as well-being

It is an important aspect of health that the body and mind are healthy, both in order and function. But we may ask for the criteria of such well-functioning. How do we know that the body and mind function well? When is the body in balance?

A traditional answer is that the person's subjective well-being is the ultimate criterion. Simply put: when a person feels well, then he or she is healthy. This statement certainly entails problems, since a person can feel well and still have a serious disease presenting. The general idea can, however, be modified to cover this case too. The individual with a serious disease will sooner or later have negative experiences such as pain, fatigue, or anguish. Thus, the ultimate criterion of a person's health is his or her present or future well-being.

It is a difficult task to characterise the well-being constituting health. If one includes too much in the concept there is a risk of identifying health with happiness. Indeed, a common accusation directed against the World Health Organisation (WHO) definition of health that it falls into this trap. Health cannot reasonably be identical with complete physical, mental, and social well-being, many critics say. The absurd conclusion of this conception could be that all people who are not completely successful in life would be deemed unhealthy.

Some authors (for instance, Hans-Georg Gadamer) have pointed out that phenomenological health (or health as experienced) tends to remain a forgotten subject. Health is in daily life hardly recognised at all by its subjects. People are reminded of their previous health first when it is being disrupted, when they experience the pain, nausea, or anguish of illness. Health is “felt” only under special circumstances, the major instance being after periods of illness when the person experiences relief in contrast to the previous suffering.

Thus, although well-being or absence of ill-being is an important trait in health, most modern positive characterisations of health have focused on other traits. One such trait is health as a condition for action, i.e. ability. I will return to this idea when I discuss holistic theories of health below.

Two contemporary streams of philosophy of health and disease

As I said, two main streams of theories of health and disease have recently appeared in the arena. One main stream is sometimes called the medical one, or the bio-statistical one. What is typical of philosophers within this stream is that they claim that the concepts of health and disease and their allies – there is a whole network of medical concepts including illness, injury, impairment, defect, disability and handicap- are, or can be treated as, biological, or in certain cases psychological, measurable concepts. “Health” and “disease” are biological concepts in the same sense as “heart” and “lung” and “blood-pressure” are biological concepts. In particular, there is, according to this position, nothing evaluative or subjective about the concepts of health and disease.

The other main stream in the philosophy of health involves a completely opposing position regarding these basic matters. According to these philosophers, who are often called normativists or holists, health and disease are intrinsically value-laden concepts. They cannot be totally defined in biological or psychological terms, if these terms are supposed to be value-neutral. To say that somebody is healthy *means* that this person is in a good state of body or mind, the holist claims. And to say that somebody has attracted a disease is to say that this person has attracted something which is bad for him or her.

What I have done so far is just to give a superficial and rough demarcation of two lines of thought within this subject. It is very complicated to spell out and disentangle the different versions of these lines of thought. At least on the holistic side there are a number of versions. What I shall do here is rather to simplify matters and concentrate on a specific version of each line of thought and analyse them in more detail.

Boorse's biostatistical theory of health and disease

The choice of theory on the biological side is very easy. The articles by the American Christopher Boorse have dominated the arena. They have also been the target of most of the normative counterclaims. In presenting Boorse's theory I shall use the most recent formulations made by Boorse himself in his long defensive article, published in 1997, titled *A Rebuttal on Health*.

The aim of Boorse's biostatistical theory of disease (*BST*) is to analyse the normal-pathological distinction. In order to capture the modern Western concept of disease Boorse proposes an explication of the ancient idea that the normal is the same as the natural in saying that health is conformity to species design. In modern terms, Boorse says "species design is the internal functional organization typical of species members, viz.: the interlocking hierarchy of functional processes, at every level from organelle to cell to tissue to organ to gross behaviour, by which organisms of a given species maintain and renew their life" (1997, page 7). All conditions which are called pathological by ordinary medicine are disrupted part-functions at some level of this hierarchy, he says.

With this general description as a background Boorse presents the following definitions:

1. The *reference class* is a natural class of organisms of uniform functional design; specifically, an age group of a sex of a species, such as the human being.
2. A *normal function* of a part or process within members of the reference class is a statistically typical contribution by it to their individual survival and reproduction.
3. A *disease* is a type of internal state which is either an impairment of normal functional ability, i.e. a reduction of one or more functional abilities below typical efficiency, or a limitation on functional ability caused by environmental agents.
4. *Health* is the absence of disease" (1997, page 7-8).¹

An action-theoretic theory of health and disease

Some of the theories on the holistic side also focus on goals, but they do so in a very different way. They do not refer to biological goals but to goals in the or-

Footnote 1: In my most recent formulations, for instance in *Action, Ability and Health*, page 93, I use the locution "accepted circumstances" instead of "standard circumstances."

dinary human sense, viz. goals of intentional actions. When we intend to do something we automatically intend to achieve a goal. Such a goal is not a goal of just a particular organ. It is a goal of the whole human being. Thus, these theories are often called holistic theories.

It is significant that the holistic theories (HTH) consider the concept of health to be the primary one and disease as a secondary concept. Health has its basis on the level of the whole person. It is the person, not the individual organs, who is healthy. Let me put this general idea of health in the old way once expressed by Galen, the famous Roman physician and philosopher from 200 AD: *Health is a state in which we neither suffer from evil nor are prevented from the functions of daily life.* Let me then introduce my own specification of this general idea: A person A is completely healthy, if and only if, A is in a mental and bodily state, given standard circumstances, which is such that A has the second-order ability to realise all his or her vital goals, i.e. the states of affairs which are necessary and together sufficient for A 's minimal happiness in the long run.

According to the HTH a person is to some extent ill when he or she does not fully possess such ability. A state of illness can have various causes within the person's body or mind. Such causes of ill health as are common or typical are what we designate as diseases. Thus, diseases, according to the HTH, are such bodily and mental states of affairs that contribute to their bearer's ill health.

Two kinds of phenomena have a central place in traditional holistic accounts of health and illness. First, the presence of a certain kind of feeling, of ease or well-being in the case of health, and of pain or suffering in the case of illness; second the phenomenon of ability or disability, the former an indication of health, the latter of illness. These two kinds of phenomena are in many ways interconnected. There is first an empirical, causal connection. A feeling of ease or well-being contributes causally to the ability of its bearer. A feeling of pain or suffering may directly cause some degree of disability. Conversely, a subject's perception of her ability or disability greatly influences her emotional state.

In my own analysis I make an assumption of a strong connection between suffering and disability, where suffering is taken to be a highly general concept covering both physical pain and mental distress. A person cannot experience great suffering without evincing some degree of disability. But the converse relation does not always hold: a person may have a disability, and even be disabled in several respects, without suffering. There are, for instance, paradigm cases of ill health where suffering is absent. One obvious case is that of coma, when a person does not feel anything at all. Another concerns certain mental disabilities and illnesses. In general, when a patient cannot reflect properly on her own situation, then her disabilities need not have suffering as a consequence. In short, therefore, wherever there is great suffering there is disability, but the converse is not true.

These observations imply that the concept of disability must have a more central place in the *defining* characterization of ill health than the corresponding concept of suffering. If one of these notions is essential to the concept of ill health it must be disability. This conclusion does not deny the extreme importance of pain and suffering – as experiences and not just as causes of disability – in most instances of ill health.

Towards an assessment of the BST and the HTH

Consider now the two theories which we are going to compare.

I. *The BST approach.* *A* is completely healthy, if and only if, all organs of *A* function normally, i.e. make their species-typical contribution to the survival of the individual and the species, given a statistically normal environment. A disease is a subnormal functioning of a bodily or mental part of the human being.

II. *The HTH approach.* *A* is completely healthy, if and only if, *A* is in a bodily and mental state which is such that *A* is able to achieve all his or her vital goals, given standard circumstances. A disease is a bodily or mental process which tends to reduce the health (as holistically understood) of the human being.

What are the criteria for assessing concepts of health, illness and disease? By what standards can we say that either of the two theories is superior to the other?

There are several possible criteria for assessing the concepts. I will here only choose two, but I think important, criteria: usefulness in medical practice and usefulness in public health contexts. Let me then first consider the *medical encounter*, the encounter between a potential patient and a medical carer (a doctor, a nurse or a paramedic). In order to do this I shall tell a short story.

1. *A person approaches health care with a problem.* John approaches his family doctor with a problem. He says that he has been ill for some time. He has had considerable pain in his stomach and this has prevented him from going to work for a week. He says that he must have some disease. He cannot explain his ill health otherwise. Here we see that John asserts that he is ill. He has not made any inspection of his body in order to establish this fact. He has noticed his pain (a pain which has no immediate external cause) and he observes that he is prevented from going to work. He assumes that there is a disease which is responsible for this problem.

2. *The doctor diagnoses the problem and treats the patient.* The doctor makes an examination of John. He tries to assess the nature of the problem and when he is convinced about its nature, he seeks the causes of it. Given his medical training he will in the first instance try to find the causes of the problem in the organic functioning of John's body. In short, he seeks some disease. It is however important here to see that he is not seeking a disease for its own sake. He is not seeking any old malady. He wants to find *the cause of the patient's problem*, primarily

in terms of the disease language to be found in medical classifications and textbooks. Having found a disease that he believes to be the cause of the problem he starts treating it *lege artis*, i.e. according to the recommendations of the contemporary art of medicine.

3. *The patient is healthy again when he or she no longer has the problem.* The medical encounter is considered successful, in particular by John, when he no longer feels the pain in the stomach and can go to work as usual.

This simple exposition of the typical successful medical encounter indicates to me that the health concept used is a variant of the HTH. The establishment of the fact that John is ill, in the first place, does not presuppose any internal inspection on the organ level. John can himself (at least equally as well as the doctor) determine that he is in a state of ill health. Ill health for John is when he is in pain and unable to do something urgent for him, viz. go to work, given that the circumstances are standard, i.e. not in themselves directly preventative.

Second, it is clear that health as assumed by the patient, as well as by the health care personnel, is a state of affairs *over and above the absence of disease*. Health has not been restored just because a disease has been cured. Normally, the patient is not completely healthy, i.e. he cannot go to work, until after a time of recovery and rehabilitation. This also speaks in favour of the HTH interpretation of health.

Consider now an example taken from the field of general health promotion. There are nowadays many health-promotive campaigns in all countries which concern things such as healthy eating, physical exercise, moderate consumption of alcohol and abstention from smoking. What are the underlying reasons for these health promotion campaigns, according to the two models of BST and HTH? Are both models equally successful in providing a philosophical foundation for this work? Let us call this question *the case of general health-promotive programmes*.

The answer to the question about which model motivates health promotion efforts best is certainly dependent on how the situation is interpreted. A protagonist of the HTH would say that this case clearly speaks in favour of the HTH. General health promotion, they would say, has not primarily to do with the prevention of disease. The primary aim is that the subject should feel hale and hearty and in general be able to achieve the things he or she is aiming for. This goal certainly presupposes the prevention of all serious diseases. It need not, however, presuppose the prevention of all pathology. Being fit and able is clearly compatible with the presence of many trivial diseases.

A defender of the BST, on the other hand, would perhaps argue along the following lines: It may be true that a general health-promotive programme need not have identified a particular disease or range of diseases as its target. From this does not follow, however, that the goal is not to prevent the incidence of serious disease. In the case where abstention from smoke or alcohol is at issue it

is clear that there are some salient diseases that the promoters have in mind. Cardiac and respiratory diseases, as well as a number of cancers, are in focus in the case of smoking. Neurological diseases, liver cirrhosis, and indeed physical accidents are in focus in the case of grave alcohol abuse. If they aim for something more they are, one could argue, also some other kind of programme, for instance, a “fitness programme,” which should be held logically separate from the health promotion proper.

My answer to this, in favour of the HTH, is the following. It seems very artificial and implausible to say that broad health-promotive programmes, with their very general recommendations concerning people’s lifestyle, are aiming at disease prevention and nothing more. To say that the remaining part of the programme is logically unrelated to health seems to be a purely theoretical stipulation against the ordinary use of language. To adopt the BST as being the most adequate theory to propel general health promotion would, I think, be to legislate against ordinary language.

This then completes my brief argument in favour of the holistic theory of health.

On the relation between health and happiness

An important question remains to be answered. What is the relation between health and happiness given a holistic approach? If health has to do with the realisation of a person’s vital goals then it seems as if health comes quite close to happiness. But is this a sign that something is wrong with the holistic theory?

Before answering this question we may notice that there exists a celebrated definition of health which comes even closer to identifying health with happiness. Let us refer again to the definition from WHO, which says that “health is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity.” This definition characterises an ideal state, a state which is almost never reached by any human in the world. It is a utopian definition, which could hardly be used in the context of ordinary health care. It is also highly debatable whether it is at all adequate to capture the notion of health.

My own attitude to conceptual analysis is that notions which are held apart in ordinary language should also be held apart in a philosophical reconstruction of these concepts. According to ordinary understanding health and happiness are different. A healthy man can be unhappy, for instance because of financial difficulties or because of the loss of a loved one. An unhealthy person can be very happy. One can easily imagine even a dying person who is happy in the presence of her family and in the firm conviction that she is soon going to meet her God.

So is there then a conceptual relation at all between health and happiness in my view? Yes, there is and I think that this is also in accordance with the ordinary understanding. It is evident that health is normally conducive to happiness. If one has the capabilities involved in health, if one can do most of the things one wants to do, then it is quite likely that one is happy with life. Conversely, if one is disabled and is in pain, then most probably one is quite unhappy. This all follows from standard definitions of happiness where happiness is understood as an emotion due to one's recognition that the conditions in life are as one would like them to be.

I do not only think that this holds as a matter of empirical fact. I also think that one should define health as a state of affairs which tends to lead to a certain degree of happiness. This is also what I do in my more technical characterisation to be found, for instance in the book *Health, Science and Ordinary Language*.

Conclusion

In this paper I have set out to discuss a set of topics related to our understanding of health. I have noticed the great interest dedicated to health by a majority of Westerners of today and tried to formulate an explanation of this fact in terms of the strong secular movement in the modern world, and also in terms of the medical development. After an introduction of historical approaches, I described two major competing conceptions of health and related concepts. I have tried to compare a biostatistical theory of health and disease with a holistic one. I have noticed the essential differences and similarities between the two approaches. I have also initiated an assessment of the two conceptions, mainly from the point of view of medical practice and public health.

My conclusions from this preliminary assessment are the following:

a) The health concept used in clinical practice is related to vital goals and not just to survival. Moreover, health is something over and above the absence of disease.

b) The health concept used in the context of general health promotion is, I argue, much more naturally interpreted along holistic lines than along biostatistical lines.

Finally, I have commented on the relation between the notions of health and happiness. I have argued for the case that these notions are separate, but still related to each other. Health is in my view a typical, but indeed not necessary, contributor to happiness.

References

- Boorse C (1997). A Rebuttal on Health. In: *What is Disease? Biomedical Ethics Reviews*. Humber J, Almeder R, editors. New Jersey: Humana Press. Pp: 1-134.
- Constitution of the World Health Organization (1946). In: *Official Records of the World Health Organization*, Vol. 2:100.
- Gadamer H-G (1993). *Über die Verborgenheit der Gesundheit*. Frankfurt am Main: Suhrkamp Verlag.
- Galen (1997). *Selected Works*, translated with an introduction and notes by PN Singer. Oxford: Oxford University Press.
- Kallenberg K, Bråkenhielm CR, Larsson G (1997). *Tro och värderingar i 90-talets Sverige*. Örebro: Libris. [In Swedish]
- Nordenfelt L (2000). *Action, Ability and Health: Essays in the Philosophy of Action and Welfare*. Dordrecht: Kluwer Academic Publishers.
- Nordenfelt L (2001). *Health, Science and Ordinary Language*. Amsterdam: Rodopi.
- Plato (1998). The Republic. In: *Complete Works*. Cooper JM, editor. Indianapolis: Hackett Publishing Company.
- Pörn I (1993). Health and Adaptedness. *Theoretical Medicine*, Vol. 14: 295-304.
- Singhal GD, Patterson TJS (1993). *Synopsis of Ayurveda*, based on a translation of the Treatise of Susruta. Delhi and Oxford: Oxford University Press.

FROM “ACTS OF GOD” TO SAFETY PROMOTION – A BRIEF HISTORY OF INJURY RESEARCH

PER NILSEN

Introduction

Injuries constitute a major public health problem. Throughout the world, approximately 5 million people die annually as a result of injuries (WHO, 2002). This figure is expected to increase to more than 8 million people in 2020. As deaths from infectious diseases have declined, injuries have become the third leading cause of overall mortality and the leading cause of death among the 1 to 40 year old age group in industrialised countries (WHO, 2004). However, fatal injuries are only part of the picture. Millions of people are injured each year and survive. For every death due to injury in Sweden, there are approximately 30 hospitalisations and 200 out-patient injuries treated at emergency departments (SRV, 2004).

For many who experience an injury, it causes temporary pain and inconvenience, but for some, an injury leads to disability, chronic pain, and a profound change in lifestyle. An injury affects more than just the individual injured; it affects everyone who is involved in the injured person's life. With a fatal injury, family, friends, co-workers, employers, and other members of the injured person's community feel the loss. In addition to experiencing grief, they may experience a loss of income and/or the loss of a primary caregiver. With a non-fatal injury, family members are often called upon to care for the injured person, which can result in stress, time away from work, and lost income. Friends of the injured may be called upon to help out and the injured person's employer may struggle with temporary or permanent replacements. Others in the community such as neighbours, volunteer groups, and religious organisations, may also feel the effects of the injury.

Although the greatest cost of injury is caused through human suffering and loss, the financial costs associated with injuries are far from trivial. Almost 50% of the world's injury-related mortality occurs in young people, aged between 15 and 44 years, who comprise the most economically productive members of the

global population (WHO, 2002). Limited societal resources are required to treat and rehabilitate injured workers. Societal costs are incurred when injuries take place, as absenteeism leads to a loss of productivity and the previously discussed intangible costs affect injured workers and their social networks.

This text provides a brief account of the evolution of injury research. While injuries have been a leading cause of death and disability throughout the history of mankind, they were not scientifically studied until well into the twentieth century. The most influential factor delaying scientific attention to injuries was probably the notion that they are unpreventable random events. William Haddon, Jr. observed in 1968 that the injury field “still includes the only substantial, remaining categories of human morbidity and mortality still viewed by most laymen and professionals alike in essentially pre-scientific terms” (Haddon, 1968, page 1431). However, the pioneering work of Haddon and other injury researchers transformed how injury was conceptualised. With evolving understanding of injury in scientific terms, opportunities for prevention of injury became possible.

Chance or fault

For centuries, injuries were regarded either as random, unpredictable, and unavoidable “acts of God” or as untoward consequences of human malevolence or carelessness (Reason, 2000). However, with the advent of industrialisation in the nineteenth century, environmental risk factors for injury became more discernible and the challenges of “accident prevention” began to receive attention (Bonnie & Gayer, 2002). The industrial revolution led to urbanisation and the development of factories. These environments produced new hazards, significantly increasing the injury risk, to the extent that injuries became an accepted part of ordinary life (Berger & Mohan, 1996).

Political movements for worker protection developed in Europe in the mid-nineteenth century and later in America, yet injuries were considered an unavoidable consequence of economic development, a perspective which is an ongoing issue in the developing countries of the world even today. Although interest in work and road environment safety grew over the course of the nineteenth century, systematic scientific inquiry was rare as injury was predominantly considered in terms of chance and fault (Stevenson *et al.*, 2004).

For much of the first half of the twentieth century, injury research focused mainly seeking to identify human errors that led to injury. The research was based on the premise that people who were injured were careless, stupid, or indifferent (AAP, 1997). In keeping with the level of knowledge of the day, interventions focused exclusively on the implementation of educational measures in order to eradicate careless behaviour (Waller, 1989).

This emphasis on individual responsibility has been attributed to the highly individualistic, independent frontier America. The focus on educating people how to avoid accidents is understandable given the historical settings of sparsely populated areas where formal societal structure was still lacking and where people made their own tools and built their own houses (Christoffel & Gallagher, 1999). Another reason for the person-oriented approach may lie in the emergence of the new field of psychiatry, which emphasised emotional antecedents to life events (Waller, 1994). The “theory of accident proneness,” devised in the 1920s, postulated that people were subconsciously “accident prone.” According to this theory, some individuals possess certain stable properties that make them particularly liable to accidents. A consequence of the theory was the notion that accidents should be prevented by selecting individuals, e.g. by using various testing procedures, and allocating them tasks that are appropriate. However, it proved difficult to distinguish accident-proneness properties of the individual from variations in exposure to hazards in the environment (Harms-Ringdahl, 1993).

Biomechanics and epidemiology

The first changes to the prevailing attitude of attributing injuries to chance or individual fault emerged in the 1940s with the work of Hugh De Haven, a pilot and physiology researcher. In 1942, De Haven published a landmark paper in *War Medicine* entitled “Mechanical Analysis of Survival in Falls from Heights of Fifty to One Hundred and Fifty Feet.” De Haven had ruptured his liver, pancreas, and gall bladder in an airplane crash in 1919. During his convalescence, he began to question the inevitability of injury as a result of aviation crashes. However, he did not ask why his plane had crashed, but rather asked why he had survived while another occupant had been killed (Winston, 2000). By investigating how people successfully survived falls of 50 to 150 feet, in some cases with only minor injury, through proper dispersion of kinetic energy in amounts as great as 200 times the force of gravity, De Haven demonstrated that damage in an injury event was not inevitable. He concluded, “It is reasonable to assume that structural provisions to reduce impact and distribute pressure can enhance survival and modify injury within wide limits in aircraft and automobile accidents” (De Haven, 1942, as reproduced in Winston, 2000, page 68).

De Haven’s observations signalled the birth of the field of biomechanics (Waller, 1989). Further work in this area was undertaken by Lieut. Col. John P. Stapp. He conducted several experiments for the US Air Force on biomechanical influences at the time of crash events, as reported in “Human Tolerance to Deceleration,” published in the *American Journal of Surgery* in 1957 (Stapp, 1957). He is known for strapping himself into a rocket sled with a shoulder harness to

test the ability of the harness to withstand energy transfer forces on rapid deceleration.

Changes in epidemiologic research in the 1940s and 1950s further developed injury research. In the beginning of the twentieth century, work in epidemiology focused on reducing the burden of infectious disease. However, the development of vaccines and antibiotics and their success in reducing infectious diseases led to a shift in emphasis in epidemiology to the investigation of chronic diseases. As researchers began to search for new areas in which to conduct research, a few settled onto injuries (Haddon *et al.*, 1964). Harvard epidemiologist John E. Gordon saw that the study of injuries had many similarities to the study of infectious diseases. Introducing the concept of injury epidemiology, Gordon suggested in his 1949 paper “The Epidemiology of Accidents” in the *American Journal of Public Health* that injuries behaved like classic infectious diseases. His studies of injury distribution patterns, according to such factors as age, place, and time, demonstrated the non-randomness of injury events (Runyan, 2003).

Gordon further argued that each injury was the product not of one cause, but of forces from the three sources of the classic triad of epidemiology, i.e. the host (the individual at risk), the agent (the available energy), and the environment (the physical, biologic, and socioeconomic context, in Gordon’s work) in which host and agent find themselves (Gordon, 1949).

In 1961, James J. Gibson, an experimental psychologist, elaborated on Gordon’s “injury as a disease” concept by observing in “The Contribution of Experimental Psychology to the Formulation of the Problem of Safety” (included in *Behavioural Approaches to Accident Research*) that there are only five agents in all injury events, namely the five forms of physical energy: thermal, radiant, chemical, electrical, and kinetic (or mechanical) energy (Gibson, 1964). Gibson suggested that injury is due to transfer of energy to the host in amounts that exceed the threshold for tissue damage (Robertson, 1998). This conceptualisation of energy as the causal agent provided the basis to view the injury event as separate from the damage to the body (Rivara, 2001).

Haddon’s contributions

The host-agent-environment model proved to be a powerful concept to aid the development of interventions that address different aspects of the injury problem. However, Gordon, Gibson, and others failed to identify the “agents” in the model; they were confused because potential agents of injury, i.e. injurious objects such as bicycles, cars, stairs, stoves, and knives, seemed unlimited (Bonnie & Guyer, 2002). This dilemma was resolved in 1963 by William Haddon, a public health physician and engineer (who trained under Gordon). Haddon recognised that what had previously been thought to be agents were in fact vehi-

cles and vectors for the five energy forms. Hence, injurious objects like bicycles, cars, stairs, stoves, and knives simply represent a multitude of vehicles and vectors for carrying the energy. Haddon also noted that in a few types of injury events, such as drowning and hypothermia, the damage is caused, not by transfer of energy to the person, but rather by interference with normal energy exchange patterns, such as breathing or maintaining body temperature (Haddon, 1963). These groundbreaking findings, which Haddon labelled “the new theory of accident causation,” were published in the paper “A Note Concerning Accident Theory and Research with Special Reference to Motor Vehicle Accidents” (included in *Annals New York Academy of Sciences*).

Haddon’s 1964 book, *Accident Research: Methods and Approaches*, co-authored with Edward Suchman and David Klein, was a milestone in developing accident research (as it was known then) as a science. The book compiled together important research contributions from more than 50 researchers, including De Haven, Stapp, Gordon, and Gibson. Haddon and colleagues urged that injury research advance from what they termed “pre-scientific” thinking to applying scientific principles to understanding injury (Haddon *et al.*, 1964). The book established the basic principles of the injury field and was a catalyst for the subsequent development of injury research (Bonnie & Guyer, 2002).

In 1968, Haddon further contributed to the understanding of injury occurrence by demonstrating in “The Changing Approach to the Epidemiology, Prevention, and Amelioration of Trauma: The Transition to Approaches Etiologically Rather than Descriptively Based” published in the *American Journal of Public Health*, that all injury involves three consecutive phases, which he termed the “crash sequence.” This sequence begins with a phase that has variously been called the pre-injury or pre-event phase, in which loss of control in management of the energy source takes place. This phase has traditionally been referred to as an “accident.” In the second phase, known as injury or event phase, the now errant energy is transferred to people and property. The nature and extent of this transfer determines whether injury occurs and its initial severity. In the third phase, the post-injury or post-event phase, efforts are undertaken to limit any ongoing damage processes, regain physiological homeostasis, and repair the damage (Haddon, 1968).

In his seminal 1970 paper “On the Escape of Tigers: an Ecological Note” in the *American Journal of Public Health*, Haddon expanded his three-phase injury sequence model by combining it with the three epidemiologic factors (host, vector/vehicle, and environment) to form a nine-cell matrix, the so-called Haddon Matrix. The paper also described 10 prevention strategies, also arranged temporally, which represent the various ways in which energy transfer can be controlled, modified, or interrupted (Haddon, 1970). Haddon argued that understanding how the temporal events of an injury and its epidemiologic components work together can help researchers and practitioners plan prevention

strategies (Christoffel & Gallagher, 1999). Haddon later refined his matrix, creating a 12-cell matrix by listing the columns as human (or host), vehicles and equipment, physical environment, and socioeconomic environment. Still later, he revised the model to consider topics other than traffic crashes (Runyan, 2003). The Haddon Matrix has proven a highly useful tool for analysing injury events and identifying factors important to their prevention. Haddon's contributions demonstrated the practical value of using theory and conceptual models to guide prevention (Robertson, 1998).

The era of legislation

The work of Haddon was monumental; he transformed how injury was conceptualised, shaping modern injury science as a distinct field. Haddon's early 1960s work, which focused on road traffic safety, markedly influenced a young Ralph Nader. He explored engineering design of automobiles and wrote *Unsafe at Any Speed: The Designed-In Dangers of the American Automobile*. The book, published in 1965, asked why thousands of Americans were being killed or injured in car accidents when the technology already existed to make cars safer (Academy of Achievement, 2006). The chief target of the book was General Motors' Corvair, a car whose faulty rear suspension system made it possible to skid violently and roll over. More generally, Nader's book documented how the auto industry in Detroit habitually subordinated safety to style and marketing concerns. The main cause of car injuries, Nader demonstrated, was not the "nut behind the wheel" so often blamed by the auto industry, but the inherent engineering and design deficiencies of the motor vehicle (Nader, 1991).

Worried about litigation challenging the Corvair's safety, General Motors hired private detectives to tail Nader in an attempt to dig up information that might discredit him; they even had women accost him in an apparent seduction scheme. Upon learning of the company's dirty tricks, Nader successfully sued the company for invasion of privacy and forced it to publicly apologise. This remarkable incident catapulted auto safety into the public spotlight. Nader became a catalyst for consumer empowerment for improved ergonomically designed safety equipment in motor vehicles (Pearn *et al.*, 2004).

Spurred by Nader's book, the US Congress enacted the Highway Safety Act, which incorporated a system of motor vehicle safety standards organised according to Haddon's pre-injury, injury, and post-injury phases (Christoffel & Gallagher, 1999). In 1966, the Congress empowered a new federal agency, the National Highway Safety Bureau (now the National Highway Traffic Safety Administration) to set motor vehicle safety standards and to award grants for research and programs promoting highway safety. Haddon became the first director of the new agency. Similar regulations were enacted with the Occupa-

tional Safety and Health Act in 1970 and the Consumer Product Safety Act in 1972 (Rivara, 2001). Internationally, safety legislation also increased from the 1970s. Improved ergonomic design for safety became enshrined in a number of government and private organisations established to define standards for building and quality control in various professions and industries (Pearn *et al.*, 2004).

Toward safety promotion

The dominant injury prevention strategy until about the 1970s was education, with interventions and programmes aimed at teaching people how to avoid accidents on the assumption that people will act in their own interest once informed of risks and benefits (Fincham, 1992). Individual error, negligence, misuse or abuse of equipment, and carelessness were viewed as the most common causes of injuries (Barry, 1975). Consequently, most research was directed toward uncovering human factors in injury aetiology (Westaby, 1974). However, the contributions of Haddon, Nader and many others helped to shift injury prevention away from the long-standing person-oriented “victim blaming” approach. Haddon and colleagues focused attention to human-environmental interactions and to application of environmental modifications when behavioural change either was unlikely or not cost-effective (Hanson *et al.*, 2004). During the 1970s, the pendulum swung in their direction and the injury field increasingly became characterised by an emphasis on environmental strategies (Bonnie, 1999). This shift in perspective generated considerable tension between those who supported environmental responses and those who still favoured behavioural perspectives to injury prevention (Christoffel & Gallagher, 1999).

With the growing recognition that neither environmental nor behavioural solutions by themselves held the complete answer to the prevention of injuries, the 1980s saw an increasing number of injury prevention programmes that combined behavioural and environmental prevention strategies, thus balancing a personal and collective responsibility for the safety problem (Waller, 1994). Behavioural perspectives were increasingly viewed as complementary rather than antagonistic to environmental perspectives. This shift reflected an increased awareness that environmental change cannot be accomplished without changing attitudes and behaviours in a target audience (Waller, 1989). There was also a growing recognition that individuals cannot be considered separately from their social context and that multiple interventions extending beyond the individual level are most effective (Hanson *et al.*, 2004).

Today, injury prevention has become an increasingly collaborative undertaking. Injury scientists and prevention practitioners need partners in order to mount successful interventions and programs. Community-based multi-strategy programmes have emerged as an important approach to injury prevention, mak-

ing it possible to deal with injury problems in less clearly delineated areas than the road traffic and work environment, including home and leisure safety (Moller, 1992). The community-based approach underpins a growing number of national and international injury prevention movements, including the World Health Organisation (WHO) Safe Community (WHO Collaborating Centre on Community Safety, 2006), the Worldwide National Safe Kids Campaign (Worldwide Safe Kids Campaign, 2005), the Canadian Safe Communities Foundation (SCF, 2006), Australian Safe Communities Foundation (ASCF, 2006), and the Safe Communities Foundation of New Zealand (SCFNZ, 2005). These programmes are often referred to as safety promotion programmes as they target structural determinants of safety and not merely individual risk factors (Svanström, 2000). Safety promotion programmes typically involve community members and local organisations in the planning and implementation of interventions (Jeffs *et al.*, 1993).

Broadening the field

The publication in 1985 of the report *Injury In America: A Continuing Public Health Problem* by the Committee on Trauma Research (established by the National Research Council (NRC) and the Institute of Medicine (IOM) of the National Academy of Sciences (NAS)) is widely considered the most important milestone in injury research since the pioneering work of Haddon in the 1960s and 1970s. The report helped to redefine the direction of injury research, setting forth the rationale for conceptualising “injury prevention and control” as a distinct field of interdisciplinary research by drawing together what had been separate strands of scientific study within the framework of public health (Bonnie & Guyer, 2002).

Injury In America: A Continuing Public Health Problem explicitly recognised that the public health paradigm (surveillance, risk factor identification, evaluation, and intervention implementation) could be usefully applied to the prevention of intentional injuries as well as unintentional injuries (Bonnie, 1999). Knowledge about intentional injuries was identified as a major gap in current research. The report observed that these types of injuries typically have been regarded as a crime problem, rather than as a health problem. The report also expressed concern over the paucity of application of promising research results in the field of rehabilitation (Waller, 1989).

Injury In America: A Continuing Public Health Problem resulted in a remarkable acceleration of injury research not only in the US, but around the world. The report recommended a major investment in injury research, commensurate with the magnitude of the problem, and proposed the creation of a centre for injury research within the US Centers for Disease Control, now the Centers for Dis-

ease Control and Prevention (Bonnie & Guyer, 2002). Substantially increased infusions of money into the field attracted new researchers and the volume of injury research expanded dramatically after the publication of the report (Rivara, 2001). Since the 1990s, the injury field has drawn the attention of a continually broadening range of disciplines, encompassing epidemiology, biomechanics, acute care, rehabilitation, psychology, criminology, economics, and other social and behavioural sciences (Bonnie & Guyer, 2002).

During the 1990s, research and program development within the injury field gave greater attention to the study of intentional injuries, reflecting a broader movement within public health embracing the cause of violence prevention (Bonnie, 1999). A 1999 report by the Committee on Injury Prevention and Control (established by the IOM), titled *Reducing the Burden of Injury: Advancing Prevention and Treatment*, endorsed the previously stated position for continued integration of all injury prevention activities, including violence prevention, within a common framework of research and program development (Bonnie & Guyer, 2002).

Reducing the Burden of Injury: Advancing Prevention and Treatment strongly argued for the displacement of the term “accident” with the term “injury.” Injury researchers and practitioners had for many years discouraged the use of “accidents” when it refers to injuries or the events that produce them. They believe it reinforces public misconception that injuries are unpredictable and unpreventable random events or “acts of God.” Moreover, not all accidents result in injuries (Christoffel & Gallagher, 1999). In 2001, all *British Medical Journal* speciality journals, including *Injury Prevention*, resolved to “ban” inappropriate use of the word “accident” in their pages. The decision, announced in an editorial in the *BMJ*, prompted an outpouring of letters to the editor. “When the dust settled, the result was a draw with equal numbers of letter writers supporting this step toward enlightenment while all others were entirely condemnatory,” editors Pless and Hagel commented in 2005 (Pless & Hagel, 2005, page 182).

In the new century, injuries remain a profoundly important public health problem, and there is growing recognition of the significance of the problem and the need for steps to address it. As with many other public health problems, the pace of prevention may seem slow, but it is steady and its momentum is growing. In fact, even the use of the dreaded “A” word does seem to be in decline.

References

- AAP (1997). *Injury Prevention and Control for Children and Youth*. Elk Grove Village: American Academy of Paediatrics.
- Academy of Achievement (2006). Ralph Nader: A Profile. Available from: <http://www.achievement.org/autodoc/page/nad0pro-1> (last accessed: 17 June 2006).
- Angus DE, Cloutier E, Albert T, *et al.* (1998). *The Economic Burden of Unintentional Injury in Canada*. Canada: Smartrisk.
- ASCF (2006). Australian Safe Communities Foundation. Available from: http://www.iccwa.org.au/Safe%20Communities/ASCF/ASCF_brochure_f inal.pdf (last accessed: 8 January 2006).
- Barry PZ (1975). Individual Versus Community Orientation in the Prevention of Injuries. *Preventive Medicine*, Vol. 4: 47-56.
- Berger LR, Mohan D (1996). *Injury Control: A Global View*. Dehli: Oxford University Press.
- Bonnie RJ (1999). *Reducing the Burden of Injury*. Washington DC: National Academy Press.
- Bonnie RJ, Guyer B (2002). Injury as a Field of Public Health: Achievements and Controversies. *Journal of Law, Medicine & Ethics*, Vol. 30: 267-280.
- Christoffel T, Gallagher SS (1999). *Injury Prevention and Public Health*. Gaithersburg: Aspen Publishers.
- De Haven H (1942). Mechanical analysis of survival in falls from heights of fifty to one hundred and fifty feet. *War Medicine*, Vol. 2: 586-596.
- Fincham S (1992). Community health promotion programs. *Social Science and Medicine*, Vol. 3: 239-249.
- Gibson JJ (1964). The contribution of experimental psychology to the formulation of the problem of safety – a brief for basic research. In: *Accident Research: Methods and Approaches*. Haddon W, Suchman EA, Klein D, editors. New York: Harper & Row. Pp: 296-303.
- Gordon JE (1949). The Epidemiology of Accidents. *American Journal of Public Health*, Vol. 39: 504-515.
- Haddon W (1963). A Note Concerning Accident Theory and Research with Special Reference to Motor Vehicle Accidents. *Annals New York Academy of Sciences*, Vol. 107: 635-646.
- Haddon W, Suchman EA, Klein D (1964). Toward a Science of Accident Research. In: *Accident Research: Methods and Approaches*. Haddon W, Suchman EA, Klein D, editors. New York: Harper & Row. Pp: 1-12.
- Haddon W (1968). The Changing Approach to the Epidemiology, Prevention, and Amelioration of Trauma: The Transition to Approaches Etiologically rather than Descriptively Based. *American Journal of Public Health*, Vol. 58 (8): 1431-1438.

- Haddon W (1970). On the Escape of Tigers: An Ecological Note. *American Journal of Public Health*, Vol. 60 (12): 2229-2234.
- Hanson D, Vardon P, Lloyd J (2004). Safe Communities: An Ecological Approach to Safety Promotion. Available from: <http://www.jcu.edu.au/school/sphtm/documents/rimnq/Paper2.pdf> (last accessed: 17 December 2004).
- Harms-Ringdahl L (1993). *Safety Analysis – Principles and Practice in Occupational Safety*. London: Elsevier Applied Science.
- Jeffs D, Booth D, Calvert D (1993). Local injury information, community participation and injury reduction. *Australian Journal of Public Health*, Vol. 17 (4): 365-372.
- Moller J (1992). Community-Based Injury Prevention. In: *Community-Based Injury Prevention*. Cowandilla: The National Safety Council of Australia. Pp: 17-23.
- Nader R (1991). *Unsafe at Any Speed: The Designed-In Dangers of the American Automobile*. Massachusetts: Knightsbridge.
- Pearn J, Nixon J, Scott I (2004). An historical perspective. In: *The Scientific Basis of Injury Prevention and Control*. McClure R, Stevenson M III, McEvoy SP, editors. East Hawthorn: IP Communications. Pp: 5-17.
- Pless IB, Hagel BE (2005). Injury prevention: a glossary of terms. *Journal of Epidemiology and Community Health*, Vol. 59: 182-185.
- Reason J (2000). Human error: models and management. *BMJ*, Vol. 320: 768-770.
- Rivara FP (2001). Introduction: The Scientific Basis for Injury Control. *Epidemiologic Reviews*, Vol. 25: 20-23.
- Robertson LS (1998). *Injury Epidemiology: Research and Control Strategies*. New York: Oxford University Press.
- Runyan CW (2003). Introduction: Back to the Future – Revisiting Haddon’s Conceptualization of Injury Epidemiology and Prevention. *Epidemiologic Reviews*, Vol. 25: 60-64.
- SCF (2006). Safety Culture: The Safe Communities Foundation. Available from: <http://www.safecommunities.ca/ohsapril.htm> (last accessed: 12 January 2006).
- SCFNZ (2005). Safe Communities Foundation NZ. Available from: <http://safecommunities.org.nz/> (last accessed: 9 December 2005)
- SRV (2004). *Ohjyckor i siffror*. Karlstad, Sweden: Räddningsverket. [In Swedish]
- Stapp JP (1957). Human Tolerance to Deceleration, *American Journal of Surgery*, Vol. 93: 739.
- Stevenson M, Ameratunga S, McClure R (2004). The rationale for prevention. In: *The Scientific Basis of Injury Prevention and Control*. McClure R, Stevenson M III, McEvoy SP, editors. East Hawthorn: IP Communications. Pp: 34-44.

- Svanström L (2000). Evidence-based injury prevention and safety promotion: state-of-the-art. In: *Injury Prevention and Control*. Mohan D, Tiwari G, editors. London: Taylor & Francis. Pp: 181-198.
- Waller JA (1989). Injury Control in Perspective. *American Journal of Public Health*, Vol. 79 (3): 272-273.
- Waller JA (1994). Reflections on a Half Century of Injury Control. *American Journal of Public Health*. Vol. 84 (4): 664-670.
- Westaby JR (1974). A bookshelf on injury control and emergency health services: injury control. *American Journal of Public Health*, Vol. 64: 394-401.
- WHO (2002). *The Injury Chart Book*. Geneva: WHO.
- WHO (2004). *Guidelines for conducting community surveys on injuries and violence*. Geneva: WHO.
- WHO Collaborating Centre on Community Safety Promotion (2006). Introduction. Available from: http://www.phs.ki.se/csp/who_introduction_en.htm (last accessed: 2 April 2006).
- Winston FK (2000). Injury classic: The beginning of injury science. *Injury Prevention*, Vol. 6: 62.
- Worldwide Safe Kids Campaign (2006). Welcome to Safe Kids Worldwide. Available from: <http://www.safekids.org> (last accessed: 14 February 2006).

THE THEORY OF COMMUNITY-BASED HEALTH AND SAFETY PROGRAMMES – A CRITICAL EXAMINATION

PER NILSEN

Introduction

Community-based programmes have become an important strategy to enhance health and safety. These programmes have their roots in the North Karelia, Stanford Five City, Minnesota Heart Health, and Pawtucket Heart Health programmes, initiated in the 1970s and 1980s to reduce high community rates of cardiovascular disease. Since then, the belief that the community-based approach is beneficial appears to have become a deeply held conviction in public health. As noted by Cheadle and colleagues (1997, page 240), “It is almost an article of faith that locating programmes in the community and involving community members in planning, implementation, and evaluation can be an effective strategy for improving population health.”

However, despite the wide application of community-based health and safety programmes during the last 30 years, there is a paucity of evaluations from which to obtain evidence regarding effectiveness of community-based health and safety programs. The present evidence from both the health promotion and injury prevention field is inconsistent, with many programmes demonstrating modest or no effects at all (Nilsen, 2006). This weak evidence has been attributed to a number of reasons, including insufficient resources, poorly implemented programmes, lack of programme reach, methodological difficulties in study design and analysis that lead to problems in demonstrating convincing results, and unrealistically high expectations of what can be achieved through these programmes (Gielen & Collins, 1993; Fisher, 1995; Feinleib, 1996; Fishbein, 1996; Klassen *et al.*, 2000; Merzel & D’Afflitti, 2003; Nilsen, 2004; Hoffmeister & Mensink, 2004; Nilsen, 2005).

This paper examines whether there are shortcomings in the theoretical underpinning of the community-based approach that could explain the lack of strong evidence of the effectiveness of health and safety programmes. Drawing upon the literature, a theory is constructed by elucidating explicit and implicit

assumptions of community-based programmes. The theory is then put to test by analysing the extent to which the assumptions are supported by empirical evidence and the extent to which the assumptions have been applied in community-based injury prevention practice.

Principles of the community-based approach

While contemporary community-based health and safety programmes do not conform rigidly to a set of predefined criteria, most community-based programmes are based on a number of key assumptions. The seven principles presented here represent important assumptions of the community-based approach. Although described as seven distinct principles, there is considerable overlap between the individual principles. Most of the principles are a matter of degree rather than all-or-none phenomena.

Community focus

The community-based approach recognises the community as a unit of identity and the appropriate focal point for health and safety programmes. The community is both the target and the catalyst for change (Israel *et al.*, 1998). This community focus is due to the realisation that humans live in, are shaped by, and in turn shape the environment in which they live (McGee, 1998). Therefore, individuals cannot be considered separately from their environment (Goodman *et al.*, 1996; Merzel & D’Afflitti, 2003).

People’s health and safety-related knowledge, attitudes, behaviours, and skills reflect their life experiences and these experiences are determined by broader institutional structures, cultural forces, and social relations within the community (McGee, 1998). This means that explanatory models centred on intra-personal determinants are of limited value to understand individuals’ health and safety; such an understanding can be achieved only if the contexts in which people live is taken into account (Richard *et al.*, 1996; Israel *et al.*, 1998).

The community is the level where many of the processes that affect people’s health and safety transpire (Hoffmeister & Mensink, 2004). Members of a community are assumed to have a sense of community, a sense of belonging to and of sharing common aspirations with the other members of the community (Steuart, 1993; Goodman *et al.*, 1998; Israel *et al.*, 1998). It has been suggested that most people yearn to be part of a larger network of relationships that give expression to their needs for intimacy, usefulness, and belonging (Sarason, 1974) and that people tend to self-segregate, i.e. interact with others like them because of shared interests, similar cultural norms, and greater empathy toward individuals who remind them of themselves (Alesina & La Ferrara, 2000).

A community can be understood both in terms of a geographical location (town, city, municipality, etc.) and a relational entity, which refers to qualities of human interaction and social ties that draw people together (Heller, 1989). The two usages of the term are not mutually exclusive and the community concept applies equally to the geographical and relational notion of community (McMillan & Chavis, 1986). However, modern society develops community around interests and skills more than around locality, implying that communities primarily are relational entities rather than geographically defined localities; what brings people together are common interests and shared values and norms around which social relationships develop (Heller, 1989; Israel *et al.*, 1998; Goodman *et al.*, 1998).

Community member participation

A key element of the community-based approach is the principle of participation, i.e. the involvement of community members in defining the health/safety problem and finding the solutions. Community member participation refers to “the social process of taking part (voluntarily) in either formal or informal activities, programmes, and/or discussions to bring about a planned change or improvement in community life, services and/or resources” (Bracht & Tsouros, 1990, page 201). The 1978 WHO Declaration of Alma Ata recognised that people must be actively involved in the process of promoting and protecting their health (Laverack, 2004). Similarly, the 1989 Manifesto for Safe Communities, generated at the First World Conference on Accident and Injury prevention, stated that “people have a right, and some would say a duty, to participate individually and collectively in the planning and implementation of their community’s safety work” (WHO, 1989, page 4-5).

Participation by community members benefits not only the community as a whole, but also the individuals who take action within the community. Participation is assumed to lead to individual empowerment, as people gain skills in assessing needs, setting priorities, and gain control over their environment (Kreuter *et al.*, 2000). The experience of involvement enhances the integrity, skills, knowledge, and experience, as well as the equality of power, for each individual who participates (Smart, 1999). Thus, the processes associated with participation are regarded as positive activities in their own right.

A number of reasons to promote community member involvement in community-based programmes have been proposed. The principle of relevance states that change will be greatest when programmes “start where the people are” (Durham, 1963, page 143) and engage community members for their knowledge of what matters to the community population (Weiss, 1995; Gielen & Sleet, 2003). This participation engenders a sense of identification and continuing responsibility for the programme, often referred to as the principle of

ownership (Carlaw *et al.*, 1984; Thompson & Kinne, 1990). Certainly, if the community actively opposes a programme, it is unlikely that the programme will have any effect and may actually produce social disruption and harm (Treno & Holder, 1997). Programme support by local opinion leaders enhances confidence in the benefits of the programme and makes it easier for individuals to accept the programme (Hoffmeister & Mensink, 2004).

Involvement by community members is a way to incorporate local values, attitudes, and symbols into the programme components and to build the layman's perspective into the programme. Community member involvement can also facilitate pre-tests of the feasibility and acceptability of new interventions, as well as provide access to local leaders, resources, and technical skills not otherwise available (Bracht & Tsouros, 1990).

Community member participation represents a bottom-up (or grassroots) approach to programme planning and decision-making. In contrast, a top-down approach involves outside agents and/or experts defining the issue, developing strategies to resolve the issue, and involving the community to assist in implementing the programme (Laverack, 2004). While a top-down approach has been effective in some areas of injury prevention, including road and workplace safety, this approach has been questioned on the grounds that it results in limited community ownership of the injury problem and solutions, which means that community members are not encouraged to think and act for themselves as they do not perceive that safety is their responsibility. A bottom-up approach is essential in dealing with less defined injury problems, in environments where regulation and enforcement is difficult to achieve (Moller, 1992). However, the dichotomy between top-down and bottom-up approaches is not as fixed as it is sometimes portrayed, as many community-based health and safety programmes tend to combine aspects of bottom-up and top-down approaches (Laverack, 2004).

Intersectoral collaboration

Collaboration among different community sectors and organisations for a common purpose is a central element of the community-based approach (Nutbeam, 1994; Merzel & D'Afflitti, 2003). Intersectoral collaborative efforts, often referred to as community coalitions, are composed of "individuals representing diverse organisations, factions, or constituencies within the community who agree to work together to achieve a common goal" (Butterfoss *et al.*, 1996, page 66).

An important rationale for intersectoral collaboration is that a great deal of that which has a direct impact on health and safety is outside the realm of the health sector (Tones & Green, 2004). Representation from multiple community sectors, organisations, groups, and key individuals is valued because of collabo-

rators' local knowledge and capacity to translate the health and safety messages into the local culture (Hawe, 1994). The community coalition can increase the credibility for the programme, as representation from different sectors enables an understanding of and a response to "true" community needs (Granner & Sharpe, 2004). Intersectoral collaboration is a way to insure local ownership and long-term maintenance of the programme (Bracht & Tsouros, 1990). It is assumed that community coalitions can achieve a vision that would not otherwise be possible to obtain as separate actors working independently (Gajda, 2004).

Injury prevention traditionally has been developed within sectors of responsibility, such as road safety and workplace safety, often with separate consideration of injuries to children and the elderly, and typically with very little cross-reference between sectors (Jefferis *et al.*, 1993). By working together, individual entities can better coordinate services across sectors and thus provide more efficient use of local resources and reduce redundancy in community services (Parker *et al.*, 1998; Gajda, 2004).

Substantial resource requirements

The challenges involved in establishing and maintaining effective community-based health and safety programmes are considerable and require substantial resource investments (Mittelmark *et al.*, 1993; Cheadle *et al.*, 1997; Turner *et al.*, 2004). The importance of identifying and building on existing community resources is recognised. Financial resources may include funding from community agencies and foundations, technical equipment, meeting space and facilities for programmatic activities (Goodman *et al.*, 1998). Financial resources can also be "expressed" as personnel resources and/or time devoted to the injury prevention work (Björns *et al.*, 1991; O'Loughlin *et al.*, 1998; Ader *et al.*, 2001).

In addition to traditional financial resources, community-based health and safety programmes also utilise intangible resources (O'Connor, 1995). Community-based health and safety programmes require large investments of human resources in the form of leadership skills and knowledge-sharing by programme collaborators (Björns, 1991; Backe, 2003). Community involvement is increased when strong collaboration exists among the different stakeholders; these relationships constitute important relational resources (or social capital) for most community-based programmes (Rifkin, 1986; Bracht & Kingsbury, 1990; Flynn, 1995; Rosén & Jansson, 2000; Petersen, 2002).

Although the community's internal resources can be seen as the raw materials for programme operation, community-based health and safety programmes may also require resources and skills available from outside of the community (Israel *et al.*, 1998). External institutions can provide technical assistance, facilitate relationships with political and funding institutions, legitimise existing ac-

tivities, and provide financial support or leverage to raise additional funds (Cheadle *et al.*, 1997).

Long-term programme view

The community-based approach recognises the importance of taking a long-term view of health/safety problems and their solution. Developing collaborative relationships with local organisations is a slow, gradual process, often requiring years for programme management to establish an environment of trust, involvement, and true understanding of local health and safety concerns (Gajda, 2004; Turner *et al.*, 2004).

Communities are more likely to commit themselves to programme development when it is not seen as a temporary project or experiment. Programmes that have a high visibility for a short period but fail to be maintained after the initial thrust create a sense of resentment for communities (Shediac-Rizkallah & Bone, 1998). Discontinued programmes could therefore pose obstacles to subsequent community mobilisation. New programmes may encounter diminished community support in communities with a history of programmes that were abruptly or inappropriately terminated. (Pluye *et al.*, 2004).

Achieving community-wide health and/or safety effects is a lengthy process because large segments of the population must be exposed to the programme (O'Loughlin *et al.*, 1998). There is often a latency period between the beginning of a programme and its effects on population health and safety. Since effects manifest typically themselves over a longer time frame, long-term programme viability is a prerequisite for meaningfully assessing effects (Rissel *et al.*, 1995; Ader *et al.*, 2001).

Multifaceted interventions

The community-based approach uses multiple interventions addressed at multiple risk factors in multiple settings and at multiple community levels (Schwarz *et al.*, 1993). Settings are the locations in which the interventions are implemented, e.g. schools, workplaces, homes, neighbourhoods, churches, and clinical settings (Richard *et al.*, 1996), while levels may range from individuals, families, small groups, and organisations to the community at large (Kubisch *et al.*, 1995).

Multifaceted interventions may be planned and implemented within the framework of numerous individual, organisational, and community-level change models and theories, including Bandura's social learning theory, social marketing theory, Rogers' innovation-diffusion theory, community stages of readiness, and various health behaviour models and theories (Puska *et al.*, 1985; Thompson & Kinne, 1990; Fincham, 1992; Goodman *et al.*, 1996; Sleet & Gielen, 2004).

This multifaceted strategy is intended to maximise the effect of the programme throughout the community by taking advantage of a synergy that is assumed to exist among different programme components (Mittelmark *et al.*, 1993; Goodman *et al.*, 1996). Earlier injury prevention approaches addressed specific types of injuries in isolation of each other and paid little attention to whether there were common risk factors between different types of injury (Moller, 1992).

With its emphasis on dynamic interplay between different community levels and settings, the community-based approach represents a social ecological perspective, placing the behaviour of individuals within a broad social context, including interpersonal relationships, neighbourhood, organisations, community, culture, and social policy (McLeroy *et al.*, 2003). Various community levels and settings are seen as a complex, nested, and interactive system (Lefebvre *et al.*, 1998). The ecological perspective incorporates a variety of concepts derived from systems theory (e.g. interdependence and homeostasis) to understand the interrelations between people and their environments (Stokols, 1992).

Population outcome

The multifaceted interventions of community-based health and safety programmes are aimed at achieving broad health and safety effects; a population outcome is the goal (Gielen & Collins, 1993; McGee, 1998). Hence, the approach directs many interventions towards the general population in the community rather than to high-risk individuals (Hoffmeister & Mensink, 2004). The population-based strategy is an attempt to control the determinants of morbidity and mortality and to lessen risk across the population (Blackburn, 1983).

It has been postulated that a population-based strategy is beneficial whenever risk is widely diffused through the whole population (Rose, 1992). However, regardless of how much benefit the population-based strategy may offer to the community as a whole, the strategy may be of little use to a given individual, with a resultant “prevention paradox” (Rose, 1981). Still, even smaller effects can be meaningful at the community level, where a modest reduction in the level of risk within a population can have a significant public health impact. In practice, many programmes combine elements of population-based and high-risk strategies in order to more effectively reach community subgroups (Mittelmark *et al.*, 1993; Merzel & D’Afflitti, 2003; Hoffmeister & Mensink, 2004).

Examining the theory

Health and safety programmes can fail for two reasons. Failure of theory occurs when programmes activate the causal mechanisms necessary to achieve the in-

tended effects but this process does not cause the desired results due to limitations in the underlying theory. Failure of implementation happens when programmes do not set the presumed causal process in motion (Weiss, 1998). This means that community-based health and safety programmes can fail either because of shortcomings inherent in the theoretical principles and/or because plausible principles are not sufficiently applied in programme practice.

Failure of theory and application

The community focus principle is based on the premise that the community is characterised by members who have a sense of community. Hence, a population may be called a community to the extent that its members have a sense of identification and emotional connection to other members of the community. However, community-based health and safety programmes overwhelmingly define community as a geographical or geopolitical unit, e.g. a town, city, municipality, or county (Elder *et al.*, 1993), which may be larger and far more diverse and heterogeneous than relationally defined communities. Research shows that community heterogeneity (e.g. in terms of ethnicity, religion, income, educational and work experience) reduces civic engagement and participation, for example measured by how people allocate their time, money, voting, and willingness to take risks to help others (Costa & Kahn, 2002).

Community mobilisation to solve health and safety problems is more likely to occur if a community sees itself as a community (Moller, 1992; Israel *et al.*, 1998). People with a strong sense of community more easily organise themselves because a common identity and a shared fate are important bases for initial group formation (Heller, 1989). A weak sense of community resulting in limited community mobilisation has been identified as an important reason for modest results in some community-based injury prevention programmes (Jeffs *et al.*, 1993; Ozanne-Smith *et al.*, 2002).

Geographical communities include people whose primary identity is based on many different factors, e.g. culture, interest, social class, ethnicity, gender or sexual orientation (Rifkin, 1986). This implies that defining the health/safety problem and finding solutions that have community-wide relevance and effects will be more difficult in geographically defined localities, as the risks and various population characteristics may vary considerably within the community. It is notable that some of the most successful community-based programmes have been implemented in Scandinavian communities, which are highly homogeneous in terms of ethnicity, culture, and socio-economic status (Nilsen, 2004). More heterogeneous communities in Australia and New Zealand have had difficulties replicating this success (Jeffs *et al.*, 1993; Coggan *et al.*, 2000; Day *et al.*, 2001).

The degree of interconnectedness among individuals is likely to be higher in smaller communities. This is a key reason why some researchers have proposed that appropriate catchment areas for community-based programmes range between 6,000 and 20,000 inhabitants (Cart Project Team, 1997). Despite this, many community-based programmes have been implemented in very large communities. For instance, the average population size of a community designated a Safe Community by the international WHO Safe Community network of community-based injury prevention programmes is 170,000 (Bourne *et al.*, 2006). The WHO Safe Communities vary greatly in size, from the small town of Os, Norway, with 2,150 inhabitants, to the large city of Dallas, USA, with 2 million people (WHO Collaborating Centre on Community Safety Promotion, 2005). The sense of community cannot be expected to be as strong in some of the largest, most diverse areas as it would be in smaller areas, where people are likely to interact frequently with each other.

The concept of community and how such an entity is defined will influence the validity of some of the other principles of the community-based approach. For example, it is highly questionable whether programmes in large cities or areas can live up to the bottom-up ideal of the community member participation principle. Moreover, it may be difficult to achieve favourable population-level results in large, heterogeneous communities considering that many community-based health promotion and injury prevention programmes lack tailored interventions to reach different segments or subgroups of the communities (Feinleib, 1996).

The population outcome principle postulates that the goals of health and safety programmes are population-level effects. Systematic reviews of community-based injury prevention programmes show that programmes narrowly targeting specific injury categories (e.g. certain injury types and/or age groups) can be highly successful if effectiveness is measured in terms of reduction of the incidence of injuries within the targeted categories (Nilsen, 2006). However, it is quite possible to obtain highly favourable results for specifically targeted injury categories without necessarily lowering the total injury incidence of the community if the target population only account for a small proportion of the injuries occurring in a community.

Failure of theory

Community member participation and intersectoral collaboration are cornerstones of the community-based approach, yet little research has been conducted to examine the relationship between community involvement and programme effectiveness (O'Neill *et al.*, 1997; Berkowitz, 2001; Granner & Sharpe, 2004). The findings thus far offer “only marginal evidence” that community involvement yields health status changes (Kreuter *et al.*, 2000, page 49) and the results

are “insufficient to make strong conclusions about the effects of partnerships on population-level outcomes” (Roussos & Fawcett, 2000, page 375).

Failure of application

Community-based programme theory emphasises the importance of an ecological perspective, with multiple interventions delivered at multiple levels and in multiple settings within the community. There is considerable evidence that multifaceted programmes are indeed more effective than narrowly focused efforts (Simons-Morten *et al.*, 1988; Moller, 1991; Elder *et al.*, 1993; Peek-Asa *et al.*, 2004; Hanson *et al.*, 2004). Still, empirical findings suggest that there is a wide variation in the degree to which community-based health and safety programmes actually apply an ecological perspective (Merzel & D’Afflitti, 2003). Single-setting or single-strategy programmes outnumber multifaceted programmes, as practitioners still prefer to target intrapersonal determinants of health rather than unhealthy aspects of people’s environments (Richard *et al.*, 1996).

Two other important features of the community-based approach are the importance of taking a long-term programme view and mobilising substantial resources to establish, deliver, and sustain programmes. Insufficient programme duration has been identified as an important factor that explains the lack of significant effectiveness of many community-based health and safety programmes (Feinleib, 1996; Merzel & D’Afflitti, 2003; Nilsen, 2004; Hoffmeister & Mensink, 2004). Unfortunately, few evaluations of community-based health and safety programmes provide explicit or detailed information on the resource use, making it difficult to determine to what extent resources influence programme effectiveness. Still, solid empirical evidence demonstrates the utmost importance of both financial and intangible resources for programme sustainability, which is a requirement for achieving programme effectiveness (Björås *et al.*, 1991; O’Loughlin *et al.*, 1998; Ader *et al.*, 2001; Doll *et al.*, 2003; Nilsen *et al.*, 2005).

Discussion

This analysis of the community-based approach to health and safety programmes suggests that the theoretical underpinning of this approach has important shortcomings, implying that many community-based health and safety programmes do not function at an optimum level. The extent to which the different principles are applied (as prescribed by theory) in programme practice can be illustrated by a two-dimensional diagram (figure 1), with an application dimension indicating the degree to which each principle is applied in practice and

a theory dimension involving an assessment of the extent to which each principle appears to be theoretically sound.

Figure 1: Theoretical plausibility and practical application of the seven principles of the community-based approach to health and safety programs

<i>Theoretical dimension</i>		
The principle may have shortcomings	The principle appears to be plausible	
<p>Community member participation Widely applied principle but there is a lack of strong evidence supporting a relationship with programme effectiveness</p> <p>Intersectoral collaboration Widely applied principle but there is a lack of strong evidence supporting a relationship with programme effectiveness</p>		<i>Application dimension</i> The principle is applied to a large extent as intended in practice The principle is not fully or widely applied as intended in practice
<p>Community focus Many programmes are implemented in geographically defined communities with a poor sense of community that yields insufficient community mobilisation and programme reach due to considerable heterogeneity</p> <p>Population outcome Many programmes are narrowly focused and may achieve high levels of effectiveness for targeted injury categories without attaining favourable community-level effects</p>	<p>Multifaceted interventions Evidence supports the plausibility of the principle but few programmes are sufficiently multifaceted</p> <p>Long-term programme view Evidence supports the plausibility of the principle but few programmes are sufficiently long-term</p> <p>Substantial resource requirements Evidence supports the plausibility of the principle but programme effectiveness is often constrained by resource constraints</p>	

Whereas the importance of multifaceted interventions, long-term programme view, and substantial resource requirements to attain effectiveness is well supported by empirical evidence, the principles of community member participation and intersectoral collaboration are somewhat less convincing, as there is a paucity of research that links programme effectiveness to this type of community involvement.

The principles of community focus and population outcome principles have theoretical limitations and tend to be applied in injury prevention practice without adherence to the principles' underlying assumptions. Programmes to a great extent define geographical or geopolitical units as communities. However, because these entities can be highly heterogeneous and be characterised by a weak sense of community, such an approach can result in insufficient community mobilisation and inadequate reach for many programmes. Many programmes are narrowly focused and may achieve high levels of effectiveness for targeted injury categories without attaining favourable community-level effects.

The findings point to the critical importance of devoting sufficient resources to mounting and running programmes. This will allow for comprehensive community assessment and facilitate application of the principles of multifaceted interventions and long-term programme perspectives in order to achieve a favourable population outcome.

However, even if programmes were more lavishly funded, the extent to which local health and safety problems can be solved mainly or merely by mobilising local efforts may be questioned. Local communities are increasingly affected by wide, far-reaching societal trends. Indeed, in a world where societies are becoming increasingly heterogeneous and populations more mobile (Tellnes, 2005), the local, geographically defined community may lose much of its decisive influence over the lives of its population.

In the contemporary world, the notion of the local community as an independent and autonomous entity that can plan and maintain itself is losing relevance (Beck, 2000). Kempny (2000, page 8) argues that "sheer places have lost their power to define the terms of human existence" and other researchers have referred to "the end of community as we know it" and "the decline of community" when discussing how the increasingly globalised conditions of contemporary society are creating new non-local, non-spatially bounded communities which ignore distance and nullify time (Bauman, 1998; Appadurai, 1999). The globalisation process poses a serious challenge to the traditional community-based approach to injury prevention and health promotion.

References

- Ader M, Berenston K, Carlsson P, Granath M, Urwitz V (2001). Quality indicators for health promotion programmes. *Health Promotion International*, Vol. 16: 187-195.
- Alesina A, La Ferrara E (2000). Participation in Heterogeneous Communities. *Quarterly Journal of Economics*, Vol. 115: 847-904.
- Appadurai A (1999). Globalization and the research imagination. *International Social Science Journal*, Vol. 160: 229-238.

- Backe S (2003). *Systematiskt kommunbaserat säkerhetsarbete*. Karlstad: Institutionen för samhällsvetenskap. [In Swedish]
- Bauman Z (1998). *Globalization: The Human Consequences*. Cambridge: Polity Press.
- Beck U (2000). *What is Globalization?* Cambridge: Polity Press.
- Berkowitz B (2001). Studying the Outcomes of Community-Based Coalitions. *American Journal of Community Psychology*, Vol. 29: 213-227.
- Bjärås G (1991). The need of leadership for motivation of participants in a community intervention programme. *Scandinavian Journal of Social Medicine*, Vol. 19: 190-198.
- Bjärås G, Haglund BJA, Rifkin S (1991). A new approach to community participation. *Health Promotion International*, Vol. 6: 199-206.
- Blackburn H (1983). Research and Demonstration Projects in Community Cardiovascular Disease Prevention. *Journal of Public Health Policy*, Vol. 4: 398-422.
- Bourne M, Nilsen P, Hudson D (2006). Towards Safer Communities? Submitted for publication.
- Bracht N, Kingsbury L (1990). Community organization principles in health promotion: A five-stage model. In: *Health Promotion at the Community Level*. Bracht N, editor. Newbury Park; Sage Publications. Pp: 66-88.
- Bracht N, Tsouros A (1990). Principles and strategies of effective community participation. *Health Promotion International*, Vol. 5: 199-208.
- Butterfoss FD, Goodman RM, Wandersman A (1996). Community Coalitions for Prevention and Health Promotion: Factors Predicting Satisfaction, Participation, and Planning. *Health Education Quarterly*, Vol: 23: 65-79.
- Carlaw RW, Mittlemark MB, Bracht N, Luepker R (1984). Organization for a community cardiovascular health program: experiences from the Minnesota heart health program. *Health Education Quarterly*, Vol. 11: 243-252.
- Cart Project Team (1997). Community Action for Health Promotion: A Review of Methods and Outcomes 1990-1995. *American Journal of Preventive Medicine*, Vol. 13: 229-239.
- Cheadle A, Beery W, Wagner E, Fawcett S, Green L, Moss D, Plough A, Wandersman A, Woods I (1997). Conference Report: Community-Based Health Promotion – State of the Art and Recommendations for the Future. *American Journal of Preventive Medicine*, Vol. 13: 240-243.
- Coggan C, Patterson P, Brewin M, Hooper R, Robinson E (2000). Evaluation of the Waitakere Community Injury Prevention Project. *Injury Prevention*, Vol. 6: 130-134.
- Costa DL, Kahn ME (2002). Civic Engagement and Community Heterogeneity: An Economist's Perspective. Available from: <http://web.mit.edu/costa/www/costa.kahn.1.4pdf.pdf> (last accessed: 8 December 2005).

- Day LM, Ozanne-Smith J, Cassell E, Li L (2001). Evaluation of the Latrobe Valley Better Health Injury Prevention Program. *Injury Prevention*, Vol. 7: 66-69.
- Doll L, Bartenfeld T, Binder S (2003). Evaluation of Interventions Designed to Prevent and Control Injuries. *Epidemiologic Review*, Vol. 25: 51-59.
- Durham A (1963). Suggested Principles of Community Development. *International Review of Community Development*, Vol. 11: 142-151.
- Elder JP, Schmid TL, Dower P, Hedlund S (1993). Community Heart Health Programs: Components, Rationale, and Strategies for Effective Interventions. *Journal of Public Health Policy*: 463-479.
- Feinleib M (1996). Editorial: New Directions for Community Intervention Studies. *American Journal of Public Health*, Vol. 86: 1696-1698.
- Fincham S (1992). Community health promotion programs. *Social Science and Medicine*, Vol. 3: 239-249.
- Fishbein M (1996). Editorial: Great Expectations, or Do We Ask Too Much from Community-Level Interventions? *American Journal of Public Health*, Vol. 86: 1075-1076.
- Fisher EB (1995). Editorial: The Results of the COMMIT Trial. *American Journal of Public Health*, Vol. 85: 159-160.
- Flynn BS (1995). Measuring community leaders' perceived ownership of health education programs: initial tests of reliability and validity. *Health Education Research*, Vol. 10: 27-36.
- Gajda R (2004). Utilizing Collaboration Theory to Evaluate Strategic Alliances. *American Journal of Evaluation*, Vol. 25: 65-77.
- Gielen AC, Collins B (1993). Community-based interventions for injury prevention. *Family Community Health*, Vol. 15: 1-11.
- Gielen AC, Sleet D (2003). Application of Behavior-Change Theories and Methods to Injury Prevention. *Epidemiologic Review*, Vol. 25: 65-76.
- Goodman RM, Wandersman A, Chinman M, Imm P, Morrissey E (1996). An Ecological Assessment of Community-Based Interventions for Prevention and Health Promotion: Approaches to Measuring Community Coalitions. *American Journal of Community Psychology*, Vol. 24: 33-61.
- Goodman RM, Speers MA, McLeroy K, Fawcett S, Kegler M, Parker E, Smith SR, Sterling TD, Wallerstein N (1998). Identifying and Defining the Dimensions of Community Capacity to Provide a Basis for Measurement. *Health Education Behaviour*, Vol. 25: 258-278.
- Granner ML, Sharpe PA (2004). Evaluating community coalitions characteristics and functioning: a summary of measurement tools. *Health Education Research*, Vol. 19: 514-532.
- Hanson D, Vardon P, Lloyd J (2004). Safe Communities: An Ecological Approach to Safety Promotion. Available from:

- <http://www.jcu.edu.au/school/sphtm/documents/rimnq/Paper2.pdf> (last accessed: 8 December 2005).
- Hawe P (1994). Capturing the meaning of “community” in community intervention evaluation: some contributions from community psychology. *Health Promotion International*, Vol. 9: 199-210.
- Heller K (1989). The Return to Community. *American Journal of Community Psychology*, Vol. 17: 1-15.
- Hoffmeister H, Mensink GBM (2004). Community-based intervention trials in developed countries. In: *Oxford Textbook of Public Health*. Detels R, McEwen J, Beaglehole R, Tanaka H, editors. Oxford: Oxford University Press. Pp: 583-598.
- Israel BA, Schulz AJ, Parker EA, Becker AB (1998). Review of Community-Based Research: Assessing Partnership Approaches to Improve Public Health. *Annual Reviews of Public Health*, Vol. 19: 173-202.
- Jeffs D, Booth D, Calvert D (1993). Local injury information, community participation and injury reduction. *Australian Journal of Public Health*, Vol. 17: 365-372.
- Klassen TP, MacKay JM, Moher D, Walker A, Jones AL (2000). Community-Based Injury Prevention Interventions. *The Future of Children*, Vol. 10: 83-110.
- Kempny M (2000). Globalization of Democracy and Conditions for Democratic Community in the Globalised World. Available from: <http://www.rci.rutgers.edu/~culdemsm/Kempny%20Paper.htm>. Last visited 20 January 2006.
- Kreuter MW, Lezin NA, Young LA (2000). Evaluating Community-Based Collaborative Mechanisms: Implications for Practitioners. *Health Promotion Practice*, Vol. 1: 49-63.
- Kubisch AC, Weiss CH, Schorr LB, Connell JP (1995). Introduction. In: *New Approaches to Evaluating Community Initiatives: Concepts, Methods, and Contexts*. Connell JP, Kubisch AC, Schorr LB, Weiss CH, editors. Washington DC: The Aspen Institute. Pp: 1-22.
- Laverack G (2004). Health Promotion in Context. In: *Health Promotion Practice: Power & Empowerment*. London: Sage Publications. Pp: 1-15.
- McGee TK (1998). The social context of responses to lead contamination in an Australian community: implications for health promotion. *Health Promotion International*, Vol. 13: 297-306.
- McLeroy KR, Norton BL, Kegler MC, et al. (2003). Community-Based Interventions. *American Journal of Public Health*, Vol. 93: 529-533.
- McMillan DW, Chavis DM (1986). Sense of Community: A Definition and Theory. *Journal of Community Psychology*, Vol. 14: 6-23.

- Merzel C, D’Afflitti J (2003). Reconsidering Community-Based Health Promotion: Promise, Performance, and Potential. *American Journal of Public Health*, Vol. 93: 557-574.
- Mittelmark MB, Hunt MK, Heath GW, Schmid TL (1993). Realistic Outcomes: Lessons from Community-Based Research and Demonstration Programs for the Prevention of Cardiovascular Diseases. *Journal of Public Health Policy*, Vol. 14: 437-462.
- Moller J (1991). Community-based Interventions: An Emerging Dimension of Injury Control Models. *Health Promotion Journal of Australia*, Vol. 1: 51-54.
- Moller J (1992). Community-Based Injury Prevention. In: *Community-Based Injury Prevention*. Cowandilla: The National Safety Council of Australia. Pp: 17-23.
- Nilsen P (2004). What Makes Community-Based Injury Prevention Work? – In Search of Evidence of Effectiveness. *Injury Prevention*, Vol. 10: 268-274.
- Nilsen P (2005). Evaluation of Community-Based Injury Prevention Programs – Methodological Issues and Challenges. *International Journal of Injury Control and Safety Promotion*, Vol. 12: 143-156.
- Nilsen P, Timpka T, Nordenfelt L, Lindqvist K (2005). Towards Improved Understanding of Injury Prevention Program Sustainability. *Safety Science*, Vol. 43: 815-833.
- Nilsen P (2006). *Opening the Black Box of Community-Based Injury Prevention Programmes – Towards Improved Understanding of Factors that Influence Programme Effectiveness*. Dissertation thesis. Linköping: Linköping University 2006.
- Nutbeam D (1994). Inter-sectoral action for health: making it work. *Health Promotion International*, Vol. 9: 143-144.
- O’Connor A (1995). Evaluating Comprehensive Community Initiatives: A View from History. In: *New Approaches to Evaluating Community Initiatives: Concepts, Methods, and Contexts*. Connell JP, Kubisch AC, Schorr LB, Weiss CH, editors. Washington DC: The Aspen Institute. Pp: 23-64.
- O’Loughlin J, Renaud L, Richard L, Gomez LS, Paradis G (1998). Correlates of the Sustainability of Community-Based Heart Health Promotion Interventions. *Preventive Medicine*, Vol. 27: 702-712.
- O’Neill MO, Lemieux V, Groleau G, Fortin JP, Lamarche PA (1997). Coalition theory as a framework for understanding and implementing intersectoral health-related interventions. *Health Promotion International*, Vol. 12: 79-87.
- Ozanne-Smith J, Day L, Stathakis V, Sherrard J (2002). Controlled evaluation of a community based injury prevention program in Australia. *Injury Prevention*, Vol. 8: 18-22.
- Parker EA, Eng E, Laraia B, Ammerman A, Dodds J, Margolis L, Cross A (1998). Coalition Building for Prevention: Lessons Learned from the North Carolina Community-Based Public Health Initiative. *Journal of Public Health Management Practice*, Vol. 4: 25-36.

- Peek-Asa C, Dean B, Kraus JF (2004). Injury control: the public health approach. In: *Oxford Textbook of Public Health*. Detels R, McEwen J, Beaglehole R, Tanaka H, editors. Oxford: Oxford University Press. Pp: 1533-1548.
- Petersen DM (2002). The Potential of Social Capital Measures in the Evaluation of Comprehensive Community-Based Health Initiatives. *American Journal of Evaluation*, Vol. 23 (1): 55-64.
- Pluye P, Potvin L, Denis J-L (2004). Making public health programs last: conceptualizing sustainability. *Evaluation and Program Planning*, Vol. 27: 121-133.
- Puska P, Nissinen A, Tuomilehto J, Salonen JT, Koskela K, McAlister A, Kotkete TE, Maccoby N, Farquhar JW (1985). The Community-Based Strategy to Prevent Coronary Heart Disease: Conclusions from the Ten Years of the North Karelia Project. *Annual Review of Public Health*, Vol. 6: 147-193.
- Richard L, Potvin L, Kishchuk N, Prlic H, Green LW (1996). Assessment of the Integration of the Ecological Approach in Health Promotion Programs. *American Journal of Public Health*, Vol. 10: 318-328.
- Rifkin SB (1986). Lessons from community participation in health programmes. *Healthy Policy and Planning*, Vol. 1: 240-249.
- Rissel C, Finnegan J, Bracht N (1995). Evaluating quality and sustainability: issues and insights from the Minnesota Heart Health Program. *Health Promotion International*, Vol. 10: 199-207.
- Rose G (1981). Strategy of prevention: lessons from cardiovascular disease. *British Medical Journal*, Vol. 282: 1847-1851.
- Rose G (1992). *The Strategy of Preventive Medicine*. Oxford: Oxford University Press.
- Rosén M, Jansson B (2000). How to act – implementing health and safety promotion in organizations. *Health Policy and Planning*, Vol. 15: 247-254.
- Roussos ST, Fawcett SB (2000). A review of collaborative partnerships as a strategy for improving community health. *Annual Review of Public Health*, Vol. 21: 369-402.
- Sarason SB (1974). *The Psychological Sense of Community: Prospects for a Community Psychology*. San Francisco: Joffrey-Bass. Pp: 130-160.
- Schwarz DF, Grisso JA, Miles C, Holmes JH, Sutton LR (1993). An Injury Prevention Program in an Urban African-American Community. *American Journal of Public Health*, Vol. 83: 675-680.
- Shediach-Rizkallah MC, Bone LR (1998). Planning for the sustainability of community-based health programs: conceptual frameworks and future directions for research, practice and policy. *Health Education Research*, Vol. 13: 87-108.
- Simons-Morton DG, Simons-Morton BG, Parcel GS, Bunker JF (1988). Influencing personal and environmental conditions for community health: A multilevel intervention model. *Family Community Health*, Vol. 11: 25-35.
- Sleet D, Gielen AC (2004). Developing injury interventions: the role of behavioural science. In: *The Scientific Basis of Injury Prevention and Control*. McClure R,

- Stevenson M, McEvoy S, editors. East Hawthorn: IP Communications. Pp: 214-232.
- Smart H (1999). Evaluating community development for health: a survey of evaluation activity across the Lothians. *Health Education Journal*, Vol. 58: 355-364.
- Steuart GW (1993). Social and Cultural Perspectives: Community Intervention and Mental Health. In: *Health Education Quarterly*. Steckler AB, Israel BA, Dawson L, Eng E, editors. (Suppl. 1): S99-S111.
- Stokols D (1992). Establishing and maintaining healthy environments: Toward a social ecology of health promotion. *American Psychologist*, Vol. 47: 6-22.
- Tellnes G (2005). President's Column: positive and negative public health effects of urbanisation. *European Journal of Public Health*, Vol. 15 (5): 552-553.
- Thompson B, Kinne S (1990). Social Change Theory: Applications to Community Health. In: *Health Promotion at the Community Level*. Bracht N, editor. Newbury Park: Sage Publications. Pp: 45-65.
- Tones K, Green J (2004). Health and Health Promotion. In: *Health Promotion*. London: Sage Publications. Pp: 8-40.
- Treno AJ, Holder HD (1997). Community mobilization: evaluation of an environmental approach to local action. *Addiction*, Vol. 92 (Suppl. 2): S173-S187.
- Turner C, McClure R, Nixon J, *et al.* (2004). Community-based programmes to prevent pedestrian injuries in children 0-14 years: a systematic review. *Injury Control and Safety Promotion*, Vol. 11: 231-237.
- Weiss CH (1995). Nothing as Practical as Good Theory: Exploring Theory-Based Evaluation for Comprehensive Community Initiatives for Children and Families. In: *New Approaches to Evaluating Community Initiatives: Concepts, Methods, and Contexts*. Connell JP, Kubisch AC, Schorr LB, Weiss CH, editors. Washington DC: The Aspen Institute. Pp: 65-92.
- Weiss CH (1998). *Evaluation*. New Jersey: Prentice Hall. Pp: 114-151.
- WHO (1989). *First World Conference on Accident and Injury Prevention, 1989 – Manifesto for Safe Communities*. Stockholm: Karolinska Institutet.
- WHO Collaborating Centre on Community Safety Promotion (2005). Safe Community network members. Available from: <http://www.phs.ki.se/csp/> (last accessed: 8 December 2005).

This chapter draws on a chapter from Per Nilsen's 2006 dissertation thesis, *Opening the Black Box of Community-Based Injury Prevention Programmes – Towards Improved Understanding of Factors that Influence Programme Effectiveness* (Linköpings Universitet) and the article "The Theory of Community-Based Health and Safety Programs: A Critical Examination," published 2006 in *Injury Prevention*, Vol. 12: 140-145.

WORKPLACE HEALTH – INFLUENCES AND INTERVENTIONS

KERSTIN EKBERG

Introduction

Adults spend about 40% of their waking hours at work. The workplace is an important setting both in affecting people's health and as an environment in which to promote health. Health in working life is not a technical, value-free process, but rather one influenced by the ideologies, beliefs and values of key actors, such as officials, workers and unions, employers, corporations and managers, experts and others (Levenstein & Woodings, 1997). Indeed, one of the defining characteristics of the workplace setting is that it brings together a variety of groups and individuals who have different agendas and priorities with regard to work and health (Naidoo & Wills, 2000). Hence, the relationship between work and health may be viewed in different ways by different actors; a conflict of interest may arise between the actors with regard to what comes first, health or productivity.

This text begins with a look at key concepts, theories and models to explain the relationship between health and work. This is followed by a discussion of important empirical findings and research concerning strategies to improve workplace health.

Workplace health influences

Workplace health may be considered an issue for the employee and/or for the organisation and management. Numerous studies have shown strong associations between physical, organisational and psycho-social working conditions and poor health and disease. Occupational health has in particular focused on toxic and physical risk factors at work and their impact on worker health. Reorganisations, down-sizing and new management roles in recent years have shown that organisational aspects and leadership are also important for health.

Occupational health researchers have studied the complex reality of workplace health with the help of work stress models which emphasise some core

factors of importance to explain job-related health and disease. The job demand-control-social support model by Karasek and Theorell (1990) has gained considerable support in empirical studies on associations between job strain (based on a combination of high demands and low decision latitude at work) and cardiovascular disease and stroke, musculoskeletal disorders, self-rated health, and sick leave. This model puts its emphasis on distinct characteristics of the workplace from an organisational perspective and does not consider interaction aspects between the individual coping abilities, socio-economic context and work characteristics.

A more recent model, the effort-reward imbalance model, emphasises both the effort and the reward structure of work (Marmot *et al.*, 1999). The model is based on the assumption that there is a reciprocal relationship between efforts and rewards at work. Health-adverse and health-promoting psychosocial work environments are defined by the interaction between a person's cognitions, emotions, and behaviours and his or her social environment (Siegrist *et al.*, 2004). Imbalance between efforts, represented by job demands and/or obligations on the employee, and rewards, as money, esteem and job security/career opportunities, may cause sustained stress reactions. It is further assumed that this process will be amplified by a worker's overcommitment, meaning excessive work-related commitment and a high need for approval. Highly overcommitted workers will experience more strain reactions to an imbalance between effort and reward (Siegrist, 1996).

Empirical studies have shown associations between cardiovascular disease, increased smoking and alcohol intake, psychosomatic health symptoms, and poor job-related well-being (van Vegchel *et al.*, 2005). Ongoing research in the Strategies for Health arena shows that organisational conditions for learning and competence development at the workplace are also associated with perceived effort-reward imbalances and with overcommitment (Ekberg, 2006) and that leadership also affects the effort-reward imbalance (Barajas & Ekberg, 2006). Employees on long-term sick leave due to stress-related mental disorders have higher overcommitment compared to employees on long-term sick leave due to musculoskeletal disorders and compared to healthy employees (Ekberg & Strindlund, 2006).

The reciprocity perspective between the individual and the workplace is also a core feature in research on organisational justice. Organisational justice is a complex concept, comprising the aspects of distributional, procedural and interactional justice. These factors have been studied separately and in combination with regard to work performance, job satisfaction, psychosomatic health, self-rated health and sick leave (Schmitt *et al.* 1999, Väänänen *et al.* 2004). Workplaces with higher levels of organisational justice may promote a mediating effect on workers' occupational strains and could therefore contribute to employee health (Elovainio *et al.*, 2002). In an ongoing study in the Strategies for

Health arena, a high degree of perceived organisational justice has been shown to be associated with better self-rated health and lower levels of burnout (Liljegrén & Ekberg, 2006).

Theories of reciprocity between the employee and the employer are related to the construct of a psychological contract. The term “psychological contract” is used to refer to a set of beliefs regarding what employees are to give and receive from their employer (Roehling *et al.*, 2001). The present trends of globalization, restructuring and downsizing in working life may invalidate the traditional contract of long-term security in return for productivity and loyalty. Long-term reciprocal contracts are generally not only based on the exchange of pay for performance, but also on psychosocial elements such as loyalty and support (Morrison & Robinson, 1997). Rapid organisational changes may breach the contract, leading to feelings of violation and offence.

In Sweden, sick leave increased dramatically during the late 1990s and the beginning of the 2000s. Occupational groups who have previously been more or less protected from occupational disorders, including middle managers and professionals, now more commonly suffer from stress-related and psychosomatic disorders and burnout syndromes. This increase in sick leave may partly be due to the extensive labour market reorganisations that have taken place in Sweden and many other developed countries during the last decade. Far-reaching and rapid changes increase the risks of workers’ loss of control, participation and social support, and also tend to make work roles less distinct (Thornhill & Saunders, 1998).

Promoting workplace health

Interventions to improve employee health and safety may, according to Murphy and Sauter (2004), be categorised into four levels: (1) legislative/policy (e.g. regulations permitting mobility from sick leave at one workplace to another job); (2) employer/organisation (e.g. organisational culture, flexibility in work schedules, healthy work organisations, and psychological contracts); (3) job/task (e.g. participation in decision making and job enrichment); and (4) individual/job interface (e.g. stress management). This taxonomy underscores the need to approach health development and return-to-work issues in a broad and system-oriented manner.

Individual/job and job/task strategies

The main goal for any organisation, public or private, is to produce services or products to attain predetermined production goals. Most employers do believe that a healthy workforce is compatible with improved performance; it is there-

fore, from an organisational perspective, worth aiming at good health and work capacity among the employees. However, employee health tends to be viewed as an individual responsibility. Workplace health promotion often focuses on lifestyle interventions. Still, unless the organisational context is also considered and modified, such strategies are likely to have a limited effect (Naidoo & Wills, 2000).

Numerous studies have shown that participation and control are among the most pertinent health promoting factors in human life. The concept of control cannot be restricted to the job task only, rather labour market conditions and employment relations should also be included within the concept, as they may influence both the actions and health of individuals at the workplace (Aronsson *et al.*, 2002). Higher worker control within the context of work organisations may be considered a way to reduce stress and job insecurity (Karasek & Theorell, 1990).

Job insecurity is a major risk factor for worker health and well-being, with stress having been shown to limit innovativeness in organisations (Lämsäsalmi & Kivimäki, 1999). Job insecurity reflects the degree to which employees perceive dimensions of their job as threatened and feel powerless to do something about it (Ashford, 1989). However, job insecurity may also appear in seemingly safe employment conditions. The individual's own evaluation of the situation as threatening is, according to Lazarus and Folkman (1984), the determinant for feelings of job insecurity. Performance appraisal processes may, among other factors, be influenced by a worker's experience of control and empowerment. Kivimäki *et al.* (2003) showed for example that those who were still employed after down-sizing are at higher risk for deterioration of health, compared to those who left and found subsequent reemployment. As shown by Wilson (1991), participation in certain situations where worker empowerment is emphasised counteracts fear of changes in workers. Also, the opportunity for employees to participate in their work setting is related to individual development, health, and well-being. However, the concept of participation can be problematic, and there are many obstacles to its application and numerous reasons why it fails to be widely disseminated (Heller *et al.*, 1998; Neuman, 1989).

Work organisation interventions may involve job design and changes in organisational practices and human resource (HR) and social policies. Changes may affect individual level interventions involving the interface between the employees and the work process. Determinants of sickness and poor health may lie both in the environment and in the worker, and also in the dynamic interface between the two. Interventions targeting either organisational changes or individual changes neglect the importance of dynamics and interaction in this interface (Murphy & Sauter, 2004). In this context, health is not viewed as a steady state condition, but rather as developing and reacting within a socio-interactional sphere. Social exchange theories suggest that people strive for re-

ciprocity in their interpersonal and organisational relationships (Leiter & Scahufeli, 1996). These theories underscore the need for mutual interactions between, for instance, management and employees, or between the employee on sick leave and stakeholders including the workplace, health care providers, and social insurance officers. Return to work must therefore address issues of communication, competence development, social support, trust, and the values that essentially comprise organisational competence and culture. Key factors to a positive health development may include innovative solutions of infrastructures for health management within and between organisations. These infrastructures should facilitate access to knowledge, promote exchange of experiences, encourage competence development and effect shared resources in terms of human resources and work. Key barriers to positive health development in a workplace include borders limiting communication and dialogue between organisations (Friesen *et al.*, 2001), as well as barriers established by regulations preventing worker mobility.

Employer/organisational strategies

Several studies have shown that the attitude of the management is a significant factor in a worker's potential for a return to work after sickness absence (Krause *et al.*, 2001). Supervisors, however, often experience a conflict of interest between their responsibility for production and their responsibility for staff well-being when an employee goes on sick leave, and they lack training and competence in these issues (Barajas *et al.*, 2002; Strindlund, 2002). The standards and attitudes of the organisation to health and work capacity seem to be important, both in the attitudes of supervisors and in the authority vested in first line managers to work with preventive and rehabilitation measures to promote worker health (Eakin, 1992). In practice, variations occur in how supervisors define their responsibilities and how they are able to exercise this, leading to wide variations in how they work to promote a return to work for sick-listed workers.

In many studies, active participation by the sick-listed person's supervisor is said to be a particularly important factor to workforce health. The attitude of the management is therefore significant and influences the potential to return to work from sick leave. A Dutch study of employees with lumbar problems showed a 21% lower return to work in employees whose sick status was not supported by their supervisors (Krause *et al.*, 2001). These supervisors, however, often seem to lack the support of more senior managers in working with issues of ill health and rehabilitation.

Supervisors frequently experience a lack of clarity about their responsibilities to assist sick-listed workers to return to work. They experience a paucity of support and authority. This lack of guidance leads to personal interpretations of

how health and sickness absence should be dealt with, in what respects managers feel themselves responsible for adapting the workplace to a sick-listed worker, and in what respects others are responsible for this task. It might simply be a case of the extent to which an organisation gives a supervisor a mandate to work with these issues. Uncertainty leads to avoidance of responsibility, or attempts to shift it to others. Middle managers do not know how to carry out workplace adaptation and therefore need the help and support of, for instance, occupational health services (Baril *et al.*, 2003). Studies that involved teaching middle managers or supervisors about pain and how workplace attitudes promote a return to work have shown that such interventions do, in fact, lead to greater returns to work and to more adaptations of workplaces (Haldorsen *et al.*, 1997). The changes that might be required at workplaces to enable the sick-listed person to return to work thus involve a learning process for everyone involved.

Workplace involvement in return-to-work interventions for those on sick leave is more effective than a workplace merely financing or supporting medical activities to treat the sick-listed worker (Ekberg & Wildhagen, 1996, Loisel *et al.*, 1997; Loisel *et al.*, 2002; Bernacki *et al.*, 2000; Krause *et al.*, 2001). Bernacki *et al.* (2000) showed how, over time, organisational learning has developed through health management programmes in some workplaces, because of the competence interchange between leadership and professionals involved in the rehabilitation processes. Return-to-work interventions following sick leave may hence be considered a learning process over time for not only the individual but also for the organisation, as new competence is developed and attitudes are altered. To facilitate such learning processes, development of health management knowledge and infrastructures at the workplace for supporting innovative “health learning” is an important issue in developing new research approaches. The importance of management and the workplace to both promote the willingness and ability of the individual to return to work was evident from a study of people on long-term sick leave in public services (Strindlund & Ekberg, 2004). Forty per cent of those on long-term sick leave said in this study that they could return to work if the situation at the workplace changed.

A key issue for return to work is time elapsed before intervention takes place. According to Wynne and McAnaney (2004), 80% of those who are absent for six weeks or more require some assistance from their workplace in order to return to work. The probability of returning to work for those who are absent between three and six months is reduced to less than 50% and to 20% for those on sick leave more than 12 months.

If an employee’s poor health or disease is viewed as being caused by non-work-related factors, some supervisors say that they are not responsible for helping to improve the employee’s chances of returning to work. In such cases, there is probably a lack of legitimacy from upper management for supervisors

working with the issue, despite the fact that, in Sweden at least, responsibility for rehabilitation applies “regardless of how work capacity has come to be reduced” (SOU, 1990). The type of ill health experienced by workers also seems to influence the attitudes of the supervisor. Mental and psychosomatic conditions are deemed harder to deal with than physical symptoms, and these problems are largely ascribed to the individual him/herself, or his/her life situation, rather than to the workplace (Strindlund, 2002).

The longer the time that has elapsed since a sick-listed worker fell ill or took sick leave, the more important psychosocial and work-related factors appear to be. However, research into the significance of these factors for a return to work is still extremely limited. The design of the rehabilitation process affects both the motivation of the participants and the results achieved (Baril *et al*, 2003).

Many studies underline the importance of the participatory nature of the rehabilitation process. In a pilot study by Medin and colleagues (2003) workers suffering pain were on sick leave for up to eight years. More than half managed to enter work-related activities after six months’ participation in a problem-solving support group. One important reason for this success was the participatory and targeted design of the rehabilitation method. The participants’ overall goal was to return to work. Each participant was to personally state how this goal was to be achieved, and then work on achieving this goal within the framework of the rehabilitation process.

Other workplace factors impact sick-listed workers’ ability to return to their jobs, including support from co-workers. Baril and colleagues (2003) describe how perceived unfair distribution of the workload between colleagues can occur if a person cannot carry out all the usual tasks because of his or her functional impairment. In one example in which it was possible to solve this problem, the group of workers was allowed to exercise influence over which tasks were to be included in the work adaptation, as well as how various tasks were to be assigned among them.

Krause *et al.* (2001) report that people on sickness absence who experienced high demands at work before the sickness absence had lower levels of return to work in both the acute and sub-acute or chronic phases of their illnesses, compared to those who had low demands at work. Low levels of worker control at work, combined with high demands, reduced return to work by a factor of two during the sub-acute or chronic phase (but not during the acute phase). A low level of support from one’s supervisor also reduced return to work. Interestingly, Krause’s studies and others show that a worker’s previous experience with back problems facilitated a return to work. This can be interpreted to mean that previous experience with health problems aids an individual’s ability to deal with new symptoms – the individual may have learned to deal with the fear of pain and also to minimise avoidant behaviour in connection with activity (Linton, 2001).

Several studies demonstrate that work environment factors and standards, and values and priorities of management are important to the success of rehabilitation measures. The literature, however, describes few examples of how work-oriented rehabilitation can be successfully implemented in the workplace. A worker's return to the job that generated one's ill health is particularly difficult for people who are on sick leave due to stress-related mental problems or who are suffering from burnout (Astvik *et al.*, 2006). One obstacle to the application of the regulations to promote worker health is thus the current general interpretation of the "work-line strategy" (see Söderberg's chapter in this anthology) to aim for return to the sick-listed workers previous workplaces. One obstacle to the application of the regulations to promote worker health is thus the current general tendency among sick-listed workers to return to their previous workplaces.

Legislative and policy strategies

The role of the workplace as an arena for active health promotion and for prevention of disease has become increasingly recognised, not only in Swedish legislation, but also in international legislation and through policy documents from the European Union, the International Labour Organization, and the World Health Organisation, amongst others.

Regulations to improve workplace environments were introduced by the Work Environment Act, enacted in Sweden in 1991. These regulations extended the responsibilities of the employer to include a responsibility for rehabilitative environments at the workplace. Work conditions were to be adapted to the capacity of the individual employee, and the employee was given the right to influence his/her own work situation and participate in changes and development work, while the employee in turn was obliged to participate in measures designed to improve the work environment.

The National Insurance Act introduced the term "work-line strategy" into Swedish social insurance for the first time in 1992. The term referred to rehabilitation measures that were important in facilitating a return to work for people who were on long-term sickness absences or who had a temporary or permanent work disability pension. The overall objective of the work-line strategy is to maintain the workforce, i.e. as many people as possible should be able to support themselves by working. The work-line strategy also means that the public sector should prioritise active measures to bring people back to work.

The emphasis in the Swedish rehabilitation process was shifted to the workplace through the transfer of greater responsibility for workplace health to the employer. This responsibility included carrying out a rehabilitation plan and undertaking rehabilitation measures at, or in association with, the workplace, with the intention of providing work for the sick-listed employee. The Social Insur-

ance Office was given the role of coordinating the rehabilitation process between the different actors involved.

A general employer's requirement was also introduced on 1 January 1992 that obliged the employer to pay sick pay for the first two weeks of each worker's sickness absence. The purpose of this was to increase the employer's responsibility for the health of employees and create incentives for employers to initiate improvements in the work environment that would promote the health of employees.

The employer's responsibility for the rehabilitation of employees is, however, unclear, leaving wide scope for interpretation. In practice, therefore, public sector authorities have increasingly assumed the day-to-day responsibility for rehabilitation. This means, in turn, that many rehabilitation measures are carried out without workplace conditions being taken into account and without the employer becoming actively involved. The rehabilitation measures thus risk becoming largely focused on the individual on sick leave, rather than on the interaction between the individual worker and her or his workplace.

Return-to-work has become increasingly popular as a concept and a practice for promoting work capacity among employees on sick leave. According to MacEachen *et al.* (2006), return-to-work is characterised by a worker's early return to the workplace, often with modified capacity, even while the worker is still undergoing treatment. This approach is largely incorporated in the Swedish legislation and in current attempts to support part-time sick leave as a first step back to work. Return-to-work has, as a consequence, become an issue for the workplace, not only in terms of health and disease management, but also from the perspectives of law and economy.

Lack of interchange within and between organisations and an essentially "locked" labour market restrains workforce opportunities for mobility. These same factors also impact return to work or mobility into other jobs for those on sick leave (Aronsson *et al.*, 2000; Liljegren & Ekberg, 2006).

Systematic reviews of the literature on workplace-based return-to-work interventions (Franché *et al.*, 2005; MacEachen *et al.*, 2006) provide strong evidence that work disability duration is significantly reduced by work accommodation offers and regular contact between the worker's health care provider and the workplace. There is moderate evidence that disability duration is reduced by interventions which include early contact between the employee and the workplace, ergonomic work site visits (for those workers who suffer from musculoskeletal disorders) and presence of a coordinator or case manager coordinating the return-to-work process. There is also some evidence that these interventions reduce costs that are associated with the duration of work disability periods. The qualitative studies also show that central to successful return-to-work arrangements are good will and trust. There are often social and communication

barriers to return to work. Intermediary actors, such as health care and the social insurance professionals, have important roles in facilitating this process.

References

- Aronsson G, Dallner M, Gustafsson K (2000). Yrkes- och arbetsplatsinlåsning. En empirisk studie av omfattning och hälsokonsekvenser. *Arbete och Hälsa*, Vol. 5: 1-24. [In Swedish]
- Aronsson G, Gustafsson K, Dallner M (2002). Work environment and health in different types of temporary jobs. *European Journal of Work and Organizational Psychology*, Vol. 11(2): 151-175.
- Ashford SJ (1989). Self-assessments in organisations: a literature review and integrative model. *Research in Organizational Behavior*, Vol. 11: 133-174.
- Astvik W, Mellner C, Aronsson G (2006). På väg. En kvalitativ studie av långtidssjukskrivning, arbete och rörlighet. *Arbete och Hälsa*, Nr 3. [In Swedish]
- Barajas J, Nordqvist C, Ekberg K (2002). Intervention elucidate intra-organisational hindrances to employee development. In: Proceedings of the 34th Annual Congress of the Nordic Ergonomics Society. Kolmården, Sweden. Pp: 67-70.
- Barajas J, Ekberg K (2006). The influence of leadership and competence development on effort-reward imbalance. Manuscript in preparation.
- Baril R, Clarke J, Friesen M, Stock S, Cole D, The Work-Ready group (2003). Management of return-to-work programs for workers with musculoskeletal disorders: a qualitative study in three Canadian provinces. *Social Science & Medicine*, Vol. 57: 2101-2114.
- Bernacki EJ, Guidera JA, Schaefer JA (2000). A facilitated early return to work program at a large urban medical center. *Journal of Occupational and Environmental Medicine*, Vol 42: 452-469.
- Eakin JM (1992). Leaving it up to the workers: sociological perspectives on the management of health and safety in small workplaces. *International Journal of Health Services*, Vol. 22: 689-704.
- Ekberg K, Wildhagen I (1996). Long-term sickness absence due to musculoskeletal disorders: the necessary interventions of work conditions. *Scandinavian Journal of Rehabilitation Medicine*, Vol. 28: 39-47.
- Ekberg K (2006). Organisational conditions for learning and development affect perceived stress. Submitted.
- Ekberg K, Strindlund L (2006). Effort-reward imbalance and overcommitment among health care employees on sick leave. Manuscript in preparation.
- Elovainio M, Kivimäki M, Vahtera J (2002) Organizational Justice: Evidence of a New Psychosocial Predictor of Health. *American Journal of Public Health*, Vol. 92:105-108.

- Franche RL, Collen K, Clarke J, Irvin E, Sinclair S, Frank J (2005). Workplace-based return-to-work interventions: a systematic review of the quantitative literature. *Journal of Occupational Rehabilitation*, Vol. 15: 607-631.
- Friesen MN, Yassi A, Cooper J (2001). Return-to-work: the importance of human interactions and organizational structures. *Work*, Vol. 17: 11-22.
- Haldorsen EMH, Jensen IB, Linton SJ, Nygren Å, Ursin H (1997). Training work supervisors for reintegration of employees treated for musculoskeletal pain. *Journal of Occupational Rehabilitation*, Vol. 7: 33-43.
- Heller F, Pusic E, Strauss G, Wilpert B (1998). *Organizational participation. Myths and reality*. Oxford: Oxford University Press.
- Karasek R, Theorell T (1990). *Healthy work: Stress, productivity and the reconstruction of working life*. New York: Basic Books.
- Kivimäki M, Vahtera J, Elovainio MJ, Virtanen M (2003). Human Costs of Organizational downsizing: Comparing health trends between leavers and stayers. *American Journal of Community Psychology*, Vol. 32: 57-67.
- Krause N, Lund T (2001). Return to work. In: *The psychology of work place safety*. Barling J, Frone M, editors. Washington DC: American Psychological Association.
- Krause N, Dasinger LK, Deegan LJ, Rudolph L, Brand RJ (2001). Psychosocial job factors and return to work after compensated low back pain injury: a disability phase-specific analysis. *American Journal of Industrial Medicine*, Vol. 40: 374-392.
- Kristensen TS (1995). The demand-control-support model: Methodological challenges for future research. *Stress Medicine*, Vol. 11 (1): 17-26.
- Lazarus R, Folkman S (1984). *Stress, appraisal and coping*. New York: Springer.
- Leiter MP, Schaufeli WB (1996). Consistency of the burnout construct across occupations. *Anxiety, Stress and Coping*, Vol. 9: 229-243.
- Levenstein C, Woodings J (editors) (1997). *Work health, and environment. Old problems new solutions*. New York: The Guildford Press.
- Liljegren M, Ekberg K (2006). "Locked in" or "forced out"? Occupational mobility as predictor of health and burnout. Submitted.
- Linton SJ (2001). Occupational psychological factors increase the risk for back pain: a systematic review. *Journal of Occupational Rehabilitation*, Vol. 11: 53-66.
- Loisel P, Abenhaim L, Durand P, Esdaile JM, Suissa S, Gosselin L, Simard R, Turcotte J, Lemaire J (1997). A population based randomised clinical trial on back pain management. *Spine*, Vol. 22: 2911-2918.
- Loisel P, Lemaire J, Poitras S, Durand M-J, Champagne F, Stock S, Diallo B, Tremblay C (2002). Cost-benefit and cost-effectiveness analysis of a disability prevention model for back pain management: a six-year follow up study. *Occupational and Environmental Medicine*, Vol. 59: 807-815.
- Lämsäalmi H, Kivimäki M (1999). Factors associated with innovative climate: What is the role of stress? *Stress Medicine*, Vol. 15: 203-213.

- MacEachen E, Clarke J, Franche R-L, Irvin E (2006). Systematic review of the qualitative literature on return to work after injury. *Scandinavian Journal of Work, Environment and Health*, Vol 31: 257-269.
- Marmot M, Siegrist J, Theorell T, Feeney A (1999). Health and the psychosocial environment at work. In: *Social determinants for health*. Marmot M, Wilkinson RG, editors. Oxford: Oxford University Press.
- Medin J, Bendtsen P, Ekberg K (2003). Health promotion and rehabilitation: a case study. *Disability and Rehabilitation*, Vol 25 (16): 908-913.
- Morrison EW, Robinson SL (1997). When employees feel betrayed: A model of how psychological contract violation develops. *Academy of Management Review*, Vol. 22: 226-256.
- Murphy LR, Sauter SL (2004). Work organization interventions: state of knowledge and future directions. *Soz. Präventivmedizin*, Vol. 49: 79-86.
- Naidoo J, Wills J (2000). *Health promotion: Foundations for Practice*. Edinburgh: Baillière Tindall.
- Neumann JE (1989). Why people don't participate in organizational change. *Research in Organizational Change and Development*, Vol. 3: 181-212.
- Roehling MV, Cavanaugh MA, Moynihan LM, Boswell WR (2001). The nature of the new employment relationship: A content analysis of the practitioner and academic literatures. *Human Resource Management*, Vol. 39: 305-320.
- Schmitt M, Dörfel M (1999). Procedural injustice at work, justice sensitivity, job satisfaction and psychosomatic well-being. *European Journal of Social Psychology*, 29, 443-453.
- Siegrist J (1996). Adverse health effects of high effort/low reward conditions. *Journal of Occupational Health Psychology*, Vol. 1: 27-41.
- Siegrist J, Starke D, Chandola T, Godin I, Marmot M, Niedhammer I, Peter R (2004). The measurement of effort-reward imbalance at work: European comparisons. *Social Science & Medicine*, Vol. 58: 1483-1499S
- Skov T, Borg V, Orhede E (1996). Psychosocial and physical risk factors for musculoskeletal disorders of the neck, shoulders, and lower back in salespeople. *Occupational and Environmental Medicine*, Vol. 53: 351-356.
- SOU (1990). *1990:49 Work and Health*. Stockholm: Fritzes.
- Strindlund L (2002) *Sjukskrivnings- och rehabiliteringsprocessen ur chefens perspektiv – en intervjustudie av första linjens chefer. Magisteruppsats i folkhälsovetenskap*. Master's thesis in public health science. Linköping: Linköping University. [In Swedish]
- Strindlund L, Ekberg K (2004). *Långtidssjukskrivna inom Landstinget i Östergötland 2002: Diagnoser, arbetsförhållanden och rehabiliteringsbehov*. IHS rapport 2004:1. Linköping: Linköping University. [In Swedish]
- Thornhill A, Saunders MNK (1998). The meanings, consequences and implications of the management of downsizing and redundancy. A review. *Personell Review*, Vol 27: 271-295.

- Van Vegchel N, de Jonge J, Bosma H, Schaufeli W (2005). Reviewing the effort-reward imbalance model: drawing up the balance of 45 empirical studies. *Social Science & Medicine*, Vol. 60: 1117-1131.
- Väänänen A, Kalimo R, Toppinen-Tanner S, Mutanen P, Peiró JM, Kivimäki M, Vahtera J (2004). Role clarity, fairness, and organizational climate as predictors of sickness absence. A prospective study in the private sector. *Scandinavian Journal of Public Health*, Vol. 32: 426-434.
- Wilson JR (1991). Participation – a framework and a foundation for ergonomics? *Journal of Occupational Psychology*, Vol. 64 (1): 67-80.
- Wynne R, McAnaney D (2004). Employment and disability: back to work strategies. European Foundation for the Improvement of Living and Working Conditions (Eurofond). Available from: www.eurofond.eu.int.

EMPOWERMENT AND HEALTH ENHANCEMENT IN WORKING LIFE – FRAMING THE CONCEPT, REVIEWING THE EVIDENCE

HANNA ARNESON

Introduction

Working life and workplaces are important determinants of health. Although the individual also can impact her/his health by lifestyle and health behaviour, most of the determinants for health derive from working conditions, e.g. employment, income and health care (Folkhälsoinstitutet, 2005). Having a job results in an income, and this income generates many assets. Work is also a source of personal growth and development. Within work individuals develop a social identity and status. Furthermore, work often provides access to social networks, outside one's family or neighbourhood. Work also stimulates a person's self-efficacy and self-esteem. In all, having a job is good for your health (Siegrist, 2005). Yet, the workplace as an arena for health promotion is a rather recent phenomenon as described in the literature (Källestål *et al.*, 2004).

Empowerment is a fundamental principle in health promotion, regardless of arena or target group. Empowerment is the process through which people gain greater control over decisions and actions affecting their health. Empowerment implies a mobilisation of individuals and groups to strengthen basic life skills and enhance influence on underlying social and economic conditions (Nutbeam, 1998).

The notion of empowerment as an essential strategy for health promotion has, unfortunately, been inadequately documented by empirical evidence in the past, at least regarding health development in working life. Few studies have examined the relationship between empowerment in working life and health. However, this field of research is into a phase of quick expansion. There are probably two primary causes for this. There is now, in contrast to earlier research climates: 1) a valid and reliable instrument to measure empowerment in working life; and, 2) a critical mass of researchers available to explore the phe-

nomenon of workplace health promotion from a holistic perspective. Consequently, this field of research is starting to elaborate and expound.

This essay aims to give a general view of the development of an empirically-based knowledge base describing empowerment in working life and its associations with health. This piece of work updates the literature review previously presented in the thesis by Arneson (2006).

Empowerment and health enhancement

The concept of empowerment was originally described by Rappaport (1981), and occurs chiefly in public health and social work (Rappaport, 1981; Rissel, 1994; Wallerstein & Bernstein, 1988). Wallerstein and Bernstein (1988) express the overall goal of empowerment as improved quality of life and social justice. It is suggested that empowerment encompasses individual goals (e.g. ability, autonomy, control, and self-efficacy) and acts as a process (Hansson & Björkman, 2005). The uniqueness of the concept of empowerment is to be found in the process leading to the goal. The process is supposed to lead to increased influence over decisions that impact on the individual's life (Nutbeam, 1998). Participation in decision-making regarding one's own activity, and self-directed activities are crucial factors in the process of empowerment. True empowerment in this sense requires individuals to make decisions for themselves and to actively participate in events that shape their lives. Empowerment processes involve the individual identifying problems, formulating vital goals and strategies for solving the problems and achieving self-determined goals (Brookings & Bolton, 2000; Swift & Lewin, 1987; Zimmerman & Warschausky, 1998). Empowerment promotes an active approach to problem solving, an increased ability to exercise control in the environment, and increased political understanding (Kaminski *et al.*, 2000). Empowerment may therefore be considered to be a developmental process at individual, organisational and societal levels.

Empowerment in working life

Empowerment has been extensively studied in the context of working life. However, the majority of research on empowerment has been conducted within the disciplines of management, business, and organisational psychology (Arneson, 2006). Potential associations between empowerment and worker health are rarely studied. A summary of the empirical research on organizational prerequisites and psychological effects of empowerment in working life will be presented below, although the focus in this essay is the associations between empowerment and health status in working life.

Empowerment in working life is often described, understood and explored from the definition expressed by Thomas and Velthouse (1990), who outlined the concept of empowerment considering the context of working life. These researchers conceptualised empowerment in working life as an active orientation to one's work role, categorised into four cognitions: meaning, competence, self-determination and impact. Meaning relates to the value of the task goal or purpose, judged in relation to the individual's own ideal. Competence concerns the skilfulness with which an individual can perform task activities. Self-determination implies the individual's experience of being in control to initiate and regulate actions. Impact is about "making a difference," the degree to which an individual can exert influence at work (Spreitzer, 1995; Spreitzer, 2005; Thomas & Velthouse, 1990). All four dimensions must be manifested to experience empowerment. In other words, if people have the ability to make decisions (i.e. self-determination), but they do not care about the decisions they can make (i.e. lack of meaning), they will not experience themselves as empowered.

The definition and theory developed by Thomas and Velthouse (1990) was later on operationalised in order to facilitate a measure of the concept empowerment. The operationalisation of the definition into a questionnaire was performed by Spreitzer (1995), who designed the Psychological Empowerment Instrument, PEI. PEI is the most commonly used instrument for measuring empowerment in the context of working life (Boudrias *et al.*, 2003).

Studies referred to in the text below all used the PEI to examine psychological empowerment. Completed studies using the PEI concentrate on either prerequisites and antecedents for empowerment at work, i.e. structural conditions, or on the psychological experience of empowerment in working life. Empirical studies on psychological empowerment have assumed that such empowerment is facilitated by organisational conditions, and will lead to effectiveness and work satisfaction. A causal relation is often implied, although the majority of the studies to date are cross-sectional, thus making conclusions concerning causality problematic.

Antecedents for empowerment in working life

The structural prerequisites to enable empowerment are found within the organisational context (Spreitzer & Doneson, 2005). The basis for this empowerment is the sharing of power between superiors and subordinates, thereby creating more democratic organisations. In other words, the responsibility and authority to control one's job should be delegated throughout the organisation. Hitherto, one longitudinal study has reported that contextual factors predict psychological empowerment. The organisational character at the workplace in terms of formal and informal power, as well as perceived access to information,

support, opportunity and resources, predicts longitudinal psychological empowerment and work satisfaction (Laschinger *et al.*, 2004). Formal job characteristics such as visible and flexible jobs that are central to the organisation’s goals facilitate psychological empowerment, as do informal jobs characterised by alliances with supervisors, peers, and subordinates. Laschinger and her colleagues (2004) call these characteristics “structural empowerment.” They have examined and established the relationship between structural empowerment and psychological empowerment in several studies (Laschinger *et al.*, 2001a; Laschinger *et al.*, 2001b; Laschinger *et al.*, 2001c; Laschinger *et al.*, 2003; Laschinger *et al.*, 2004; Kluska *et al.*, 2004). Leadership characterised by strong socio-political support, a workplace climate with participative decision-making, access to resources and information, and good communication with the supervisor, are qualities that are also positively associated with psychological empowerment (Chan, 2004; Matthews *et al.*, 2003; Siegall & Gardner, 2000; Spreitzer, 1996; Spreitzer, 2005). Conflict with supervisors hinders empowered employees from developing or maintaining organisational commitment (Janssen, 2004). The above-mentioned associations between organizational associations with empowerment are summarised in table 1.

	Study design	
	Cross-sectional	Longitudinal
Facilitators		
Participative decision-making	√	
Strong socio-political support	√	
Good communication with the supervisor	√	
Formal power	√	√
Informal power	√	√
Access to information	√	√
Access to support	√	√
Access to opportunities	√	√
Access to resources	√	√
Hinders		
Conflicts with supervisors	√	

Table 1: Characteristics at the workplace and associations with empowerment in working life measured by the Psychological Empowerment Scale, a summary of significant findings (p<0.05).

Psychological effects of empowerment in working life

The empirical research on the psychological effects of and associations with psychological empowerment, is presented below, and summarised in table 2 as consequences of psychological empowerment. Psychological empowerment is positively associated with job satisfaction, effectiveness, innovativeness and im-

proved performance on individual and group levels, as well as improved organisational commitment (Chan, 2004; Konczak *et al.*, 2000; Sibert *et al.*, 2004; Spreitzer 1995b). It is also more commonly found in active jobs, as measured by the demand-control model by Karasek and Theorell (1990). Laschinger *et al.* (2001a; 2001b; 2001c) showed that nurses who recognised themselves as having active jobs (high control and high demands) were significantly more empowered than nurses in the high-strain group (low control and high demands). A recent study has examined the relationship between psychological empowerment and effort-reward imbalance at work (ERI) (Kluska *et al.*, 2004). The rather small study (112 nurses) found that the variance in ERI could not be explained by psychological empowerment, once the effect of structural empowerment was accounted for. There are no longitudinal studies reported on psychological effects of empowerment in working life. Consequently, no causal conclusions can be drawn.

	Study design	
	Cross-sectional	Longitudinal
Positive associations		
Job satisfaction	√	
Effectiveness in employees and managers	√	
Innovativeness	√	
Improved performance on individual and group level	√	
Improved organisational commitment	√	
Active jobs (the demand-control model)	√	

Table 2: Psychological associations with empowerment in working life measured by the Psychological Empowerment Scale, a summary of significant findings (p<0.05).

Associations with health

Although the literature about empowerment and (workplace) health promotion is based on the assumption that empowerment leads to health, only a few studies have examined this relationship. To my knowledge, there are seven studies concerning psychological empowerment and (ill) health. These studies are presented below and summarised in table 3 with respect to design characteristics. The results of these studies are summarised in table 4.

The results from these studies are concordant: empowerment in working life is recurrently associated with (ill) health in a positive direction, no matter which measures are used to operationalise health (table 4). A higher degree of empowerment is associated with better health, while lower empowerment is associated with poorer health. The results indicate the same trend measured with SF-36, EQ-5D, General health Questionnaire (GHQ), or strain symptoms (Arneson *et*

al., 2006a; Arneson *et al.*, 2006b; Arneson *et al.*, 2006c; Hochwalder & Brucefors, 2005; Spreitzer *et al.*, 1997). Empowerment is also negatively associated with self-reported sick leave during the two previous years; the higher degree of empowerment, the fewer sick leave days (Hochwalder & Brucefors, 2005).

Author	Study design	Number of participants	Context	Response rate	Instruments to measure (ill) health	Methods of analysis
Arneson <i>et al.</i> , 2006a	Cross-sectional	1246	Public sector	81 %	EQ-5D, SF-36, Copenhagen Burnout Inventory	Multiple linear regressions.
Arneson <i>et al.</i> , 2006b	Longitudinal, 2 years	1064	Public sector	65 %	EQ-5D, SF-36, Copenhagen Burnout Inventory	Multiple linear regressions.
Arneson <i>et al.</i> , 2006c	Longitudinal, 2 years	1064	Public sector	65 %	EQ-5D, SF-36, Copenhagen Burnout Inventory	Multiple linear regressions.
Ben-Zur & Yagil, 2005	Cross-sectional	228	Service sector and civil servants	100 % (volunteers)	Maslachs Burnout Inventory	Correlations. Structural Equation Modeling (SEM).
Hochwalder & Brucefors, 2005	Cross-sectional	2011	Nurses	58 %	General Health Questionnaire, Self-reported sick leave, Maslachs Burnout Inventory	Multiple linear regressions.
Laschinger <i>et al.</i> , 2003	Longitudinal, 3 years	239	Nurses	58 %	Emotional exhaustion (Maslach Burnout Inventory)	Structural Equation Modeling (SEM).
Spreitzer <i>et al.</i> , 1997	Cross-sectional	393 + 125	Middle managers + employees from an insurance company	100 % (volunteers) + 95 %	Strain symptoms (depression, anxiety, somatic symptoms), Stress	Regression Analysis.

Table 3: Characteristics of the studies that are focusing on associations between empowerment in working life (as measured by Psychological Empowerment Instrument, PEI) and health.

	Study design	
	Cross-sectional	Longitudinal
Self-rated health		
<i>SF-36</i>		
Physical function	1	
Physical role function	1	2, 3
Bodily pain	1	2, 3
General health	1	3
Vitality	1	3
Social function	1	3
Emotional role function	1	3
Mental health	1	2, 3
<i>EQ-5D</i>		
EQ-5D index	1	3
EQ-5D VAS	1	2, 3
<i>General health Questionnaire</i>		
Somatic symptoms	5	
Anxiety and insomnia	5	
Social dysfunction	5	
Depression	5	
<i>Strain symptoms</i>		
Depression	7	
Anxiety	7	
Somatic symptoms	7	
Burnout		
<i>Maslach Burnout Inventory</i>		
Emotional exhaustion	4, 5	6
Depersonalization	4, 5	
Personal accomplishment	4, 5	
<i>Copenhagen Burnout Inventory</i>		
Personal burnout	1	
Work related burnout	1	3
Client related burnout	1	
Sick leave	5	
Job stress	7	

Table 4: Reported associations ($p < 0.05$) between empowerment in working life (as measured by Psychological Empowerment Instrument, PEI) and health.

1=Arneson et al. (2006a), 2=Arneson et al. (2006b), 3=Arneson et al. (2006c), 4=Ben-Zur & Yagil (2005), 5=Hochwalder & Brucefors (2005), 6=Laschinger et al. (2003), 7=Spreitzer et al. (1997).

The findings are also unambiguous concerning burnout. The higher the degree of empowerment, the lower the degree of burnout measured both with Maslachs Burnout Inventory and Copenhagen Burnout Inventory (Arneson et

al., 2006a; Arneson *et al.*, 2006b; Arneson *et al.*, 2006c; Ben-Zur & Yagil, 2005; Hochwalder & Brucefors, 2005; Laschinger *et al.*, 2003).

However, the quality of the reviewed studies is weak, as the majority is studies of cross-sectional design (table 3). Empowerment in working life has been studied longitudinal in only two study populations. The effects of empowerment on long-term health are therefore unclear.

There are two longitudinal studies on measuring self-rated health, as this is associated with concepts related to empowerment (Arneson *et al.*, 2006b; Arneson *et al.*, 2006c). Both these studies emanate from the same study population, employees from the public sector in Sweden. One of these studies (Arneson *et al.*, 2006b) concentrates solely on associations between empowerment and self-rated health, while the other (Arneson *et al.*, 2006c) examines a combination of empowerment and social support in working life on self-rated health.

In the study focusing solely on associations between empowerment and self-rated health (Arneson *et al.*, 2006b) it was found that a higher degree of empowerment at baseline was associated with better physical role function and mental health, and less bodily pain after two years. These findings were valid for women. Men with a higher degree of empowerment at baseline reported better self-rated health measured by EQ-5D-VAS. Gender differences have not been examined or reported from any of the other researchers at this point.

In the study that examined a combination of empowerment and social support in working life and the associations with self-rated health gender differences between men and women further increased (Arneson *et al.*, 2006c). For women a combination of psychological empowerment and social support at baseline was associated with less bodily pain and better general health, vitality, social functioning, emotional role functioning, and mental health, as measured in the SF-36, as well as better self-rated health as measured by the VAS and the index in the EQ-5D after two years of employment. For men a combination of psychological empowerment and social support at baseline was significantly associated with better self-rated health as measured by EQ-5D VAS after two years.

There are three longitudinal studies investigating the associations between empowerment in working life and burnout. The results in these three studies are somewhat inconsistent. Laschinger *et al.* (2003) researched psychological empowerment as a predictor of longitudinal levels of burnout. They found that a lower degree of psychological empowerment among nurses resulted in a higher degree of emotional exhaustion three years later. This longitudinal association between empowerment and burnout could not be confirmed in a separate study conducted by Arneson *et al.* (2006b). However, an association towards work-related burnout was found by Arneson *et al.* (2006c) when a combination of empowerment and social support in working life was followed up after a two-year period.

Conclusion and implications for future research

In summary, the results of current research show that: 1) psychological empowerment is associated with self-rated health, sick leave and burnout; 2) psychological empowerment is reported to be associated with somatic and psychosocial health after a period of two years, and with burnout after three years; and 3) a combination of psychological empowerment and social support impacts self-rated health and burnout after two years more extensively than psychological empowerment does on its own. The assumption that empowerment is a fundamental principle in health promotion and an essential strategy when trying to enhance health in working life seems slowly to be supported by empirical data. Also, without a doubt, it is evident that further studies are needed.

There is a great demand for future studies that examine the relation between empowerment and health, both empirically and theoretically. Possible questions to rise are: Will new studies conducted in different work settings demonstrate that empowerment affects and predicts health? What is the causal relationship between empowerment and health? Are there differences between men and women? How does empowerment interact with social support? How long does an individual need to experience empowerment in order to develop better health? In addition, better knowledge of the antecedents to empowerment will enable interventions directed to the aetiology, thereby possibly improving empowerment as a mediating outcome towards health.

References

- Arneson H (2006). *Empowerment and health promotion in working life*. Dissertation thesis. Linköping, Sweden, Linköping University.
- Arneson H, Ekberg K (2005). Evaluation of empowerment processes in a workplace health promotion intervention based on learning in Sweden. *Health Promotion International*, Vol. 20 (4): 351-359.
- Arneson H, Ekberg K (2006). Measuring Empowerment in Working Life: a Review. *WORK: A Journal of Prevention, Assessment & Rehabilitation*, Vol. 26 (1): 37-46.
- Arneson H, Liljegren M, Nordlund A, Ekberg K (2006a). Psychological Empowerment and Self-Rated Health and Burnout in a Sample of Swedish Employees. Submitted.
- Arneson H, Liljegren M, Nordlund A, Ekberg K (2006b). Psychological empowerment at the workplace and self-rated health and burnout: a 2-year longitudinal analysis in a sample of Swedish employees. Submitted.
- Arneson H, Liljegren M, Nordlund A, Ekberg K (2006c). Psychological empowerment, social support at the workplace and self-rated health and burn-

- out: a 2-year longitudinal analysis in a sample of Swedish employees. Submitted.
- Ben-Zur H, Yagil D (2005). The relationship between empowerment, aggressive behaviours of customers, coping, and burnout. *European Journal of Work and Organizational Psychology*, Vol. 14 (1): 81-99.
- Boudrias J-S, Gobert P, Savoie A, Vandenberghe C (2003). Psychological empowerment at work: What do we know after a decade of research? *Revue Quebécoise de Psychologie*, Vol. 24 (2): 43-73.
- Brookings JB, Bolton B (2000). Confirmatory factor analysis of a measure of intrapersonal empowerment. *Rehabilitation Psychology*, Vol. 45 (3): 292-298.
- Chan YH (2004). A nomological network approach to the study of antecedents, moderator, mediators and outcomes of psychological empowerment. Dissertation Abstracts International Section A. *Humanities & Social Sciences*, Vol. 64 (10-A): 3749.
- Folkhälsoinstitutet (2005). *Folkhälsopolitisk rapport 2005*. Stockholm, Folkhälsoinstitutet. [In Swedish]
- Hansson L, Björkman T (2005). Empowerment in people with a mental illness: reliability and validity of the Swedish version of an empowerment scale. *Scandinavian Journal of Caring Sciences*, Vol. 19 (1): 32-38.
- Hochwälder J, Brucefors AB (2005). Psychological empowerment at the workplace as a predictor of ill health. *Personality and Individual Differences*, Vol. 39 (7): 1237-1248.
- Janssen O (2004). The barrier effect of conflict with superiors in the relationship between employee empowerment and organizational commitment. *Work & Stress*, Vol. 18 (1): 56-65.
- Kaminski M, Kaufman JS, Graubarth R, Robins TG (2000). How do people become empowered? A case study of union activists. *Human Relations*, Vol. 53 (10): 1357-1383.
- Karasek R, Theorell T (1990). *Healthy Work – Stress, Productivity, and the Reconstruction of Working Life*. New York, Basic Books.
- Kluska KM, Laschinger HKS, Shamian J, Wilk P (2004). Staff Nurse Empowerment and Effort-Reward Imbalance. *Nursing Leadership*, Vol. 17 (1): 112-128.
- Konczak LJ, Stelly DJ, Trusty ML (2000). Defining and Measuring Empowering Leader Behaviors: Development of an Upward Feedback Instrument. *Educational and Psychological Measurement*, Vol. 60 (2): 301-313.
- Källestål C (editor), Bjurvald M, Menckel E, Schaerström A, Schelp L, Unge C (2004). *Workplace health promotion - Effects of interventions referred to in systematic knowledge reviews and in Swedish reports*. Stockholm, National Institute of Public Health.

- Laschinger H, Finegan J, Schamian J. (2001a). Promoting nurses health: effect of empowerment on job strain and work satisfaction. *Nursing Economics*, Vol. 19 (2): 42-52.
- Laschinger HKS, Finegan J, Schamian J, Wilk P (2001b). Impact of structural and psychological empowerment on job strain in nursing work settings expanding Kanter's Model. *Journal of Nursing Administration*, Vol. 31 (5): 260-272.
- Laschinger HKS, Finegan J, Schamian J, Almost J (2001c). Testing Karasek's Demands-Control Model in Restructured Healthcare Settings. *JONA*, Vol. 31 (5): 233-243.
- Laschinger HKS, Finegan J, Schamian J, Wilk P (2003). Workplace Empowerment as a Predictor of Nurse Burnout in Restructured Healthcare Settings. *Longwoods Review*, Vol. 1 (3): 2-11.
- Laschinger HKS, Finegan JE, Schamina J, Wilk P (2004). A longitudinal analysis of the impact of workplace empowerment on work satisfaction. *Journal of Organizational Behavior*, Vol. 25 (4): 527-545.
- Matthews RA, Diaz WM, Cole SG (2003). The organizational empowerment scale. *Journal: Personnel Review*, Vol. 32 (3): 297-318.
- Nutbeam D (1998). Health promotion glossary. *Health Promotion International*, Vol. 13 (4): 349-364.
- Rappaport J (1981) In praise of paradox: A social policy of empowerment over prevention. *American Journal of Community Psychology*, Vol. 9: 2-25.
- Rissel C (1994). Empowerment: The holy grail of health promotion? *Health Promotion International*, Vol. 9 (1): 39-47.
- Siegal M, Gardner S (2000). Contextual factors of psychological empowerment. *Personell Review*, Vol. 29 (6): 703-722.
- Spreitzer GM (1995). Psychological empowerment in the workplace: dimensions, measurement, and validation. *Academy of Management Journal*, Vol. 38 (5): 1442-1465.
- Spreitzer GM (1995). An empirical test of a comprehensive model of intrapersonal empowerment in the workplace. *American Journal of Community Psychology*, Vol. 23 (5): 601-629.
- Spreitzer GM (1996). Social structural characteristics of psychological empowerment. *Academy of Management Journal*, Vol. 39 (2): 483-504.
- Spreitzer GM, Kizilos MA, Nason SW (1997). A dimensional analysis of the relationship between psychological empowerment and effectiveness, satisfaction and strain. *Journal of Management*, Vol. 23 (5): 679-704.
- Spreitzer G (2005). Empowerment at work. Available from: <http://webuser.bus.umich.edu/spreitze/Empowerment%20at%20Work.pdf>

- Spreitzer GM, Doneson D (2005). Musings on the Past and Future of Employee Empowerment. In: *Handbook of Organizational Development*, editor Cummings T. Thousand Oaks: Sage.
- Thomas KW, Velthouse BA (1990). Cognitive Elements of Empowerment: An “Interpretive” Model of Intrinsic Task Motivation. *Academy of Management Review*, Vol. 15 (4): 666-681.
- Wallerstein N, Bernstein E (1988). Empowerment education: Freire’s ideas adapted to health education. *Health Education Quarterly*, Vol. 15(4): 379-394.
- Zimmerman MA, Warschausky S (1998). Empowerment theory for rehabilitation research: and methodological issues. *Rehabilitation Psychology*, Vol. 43 (1): 3-16.

SOCIETY-BASED STRATEGIES AND MANAGEMENT ROUTINES AIMED AT PROMOTING RETURN TO WORK

ELSY SÖDERBERG

Introduction

Promoting return to work (RTW) for people on sick leave involves the expertise of numerous professionals working across different disciplines. In Sweden and many other nations, the health and medical service are responsible for medical rehabilitation, which includes – besides regular health care – advisory services, support measures and functional assessments. The aim of rehabilitation treatments provided by health and medical services is, as far as possible, to restore health and functional capacity to the individual (Jönsson *et al.*, 2004). In Sweden, the role of the social insurance authority is to assess the right to sickness benefits, and to coordinate all RTW measures of sick-listed individuals entitled to sickness benefits, using policy guidelines set by the National Insurance Act (AFL), (Söderberg, 2005). Employers bear primary responsibility for detecting and investigating the need for RTW measures, ensuring that measures are implemented, and financing these measures (SOU, 2000; Goine, 2006).

Society-based strategies for handling RTW-promoting measures constitute a complex process (Buys & Rennie, 2001; Friesen *et al.*, 2001) that involves a continuum of ongoing intentions, actions, and relations between actors (e.g. the physician, the case officer, the employer) who are affiliated with different organisations. From the viewpoint of the state, interaction and cooperation between actors is a means of lowering the rates of sickness and reducing costs to the sickness insurance system. Delayed RTW of sick-listed individuals has, however, become a critical social problem in many societies, significantly affecting economic and individual well-being (Wadell *et al.*, 2002).

An important aspect of RTW programme efficiency depends on how coordination and management routines between actors take place. To what extent are the RTW promoting measures really enhancing or do they actually hinder the possibilities to return to worklife?

Perspectives on sickness benefits in Western Europe

In most industrialised nations, including Sweden, a person is provided sickness benefits only if she or he has a disease that results in reduced capacity to work (Himmel *et al.*, 1995; Cassis *et al.*, 1996). In several countries, including the United Kingdom, Norway, and Sweden, sickness certificates are primarily issued by physicians, usually after an initial period of patient self-certification. Many European countries, including Norway, Denmark, Germany, the United Kingdom, and the Netherlands, have unambiguously stipulated maximum lengths for an individual sick leave. On the contrary, in Sweden there is, thus far, no maximum length of a sick leave period. This can lead to marginalisation of individuals and to economic difficulty in society in general. Furthermore, the number of days during which a sick leave spell can be self-certified varies between countries; for instance, a sickness certificate is required from day 4 in Germany, Denmark, and Norway, but from day 1 in France, day 8 in Sweden (Socialdepartementet, 2003), and day 8 in the United Kingdom (Shiels *et al.*, 2004). The possibility for part-time sick leave is in use in Norway, Sweden, the Netherlands, France and Denmark. In the three latter countries, the part-time sick leave is used only in relation to measures aimed at promoting RTW.

The increase in long-term sickness absence has in most Western countries been met with multiple strategies to intensify the use of RTW measures (Lierop & Nijhuis, 2000; Buys & Rennie, 2001; Friesen *et al.*, 2001). Such measures include social or vocational interventions aimed at getting an individual on sick leave back to work. The Scandinavian strategy for RTW emphasises broad coverage of sickness treatments, high benefits, and extensive re-integration efforts before disability pension is granted. It has been pointed out that re-integration represents a drawn out process of gate-keeping (Lindqvist, 2000). This can be contrasted to benefit systems with broad coverage but low benefit levels. In such systems extensive gate-keeping is not vital to maintain cost-control since benefits are low, gate-keeping to a large extent takes care of itself. The high benefit/tough gate-keeping approach can also be contrasted to benefit systems where generous benefits are reserved for core workers, while other population segments are served by low and/or means-tested benefits (Overbye, 2005).

During the 1990s, many countries carried out reforms in attempts to reduce costs related to sickness absence and to achieve equal and fair use of sickness insurance resources (Aarts *et al.*, 1996; Brage *et al.*, 1996). This has often included restrictions on the medical reasons for granting sickness benefits, rehabilitation allowances, or disability pensions. Studies (Getz & Westin, 1994; Englund *et al.*, 2000) have investigated such restrictions in insurance legislation in terms of effects on sickness certification practices or physicians, or the attitudes of physicians toward the changes in rules. The majority of physicians surveyed in these research projects have stated that if sickness absence and disabil-

ity pensions were to be certified solely on medical grounds, there would be a greater number of consultations and referrals, more acceptable diagnoses, and an increase in other types of public benefits.

Limited knowledge is available regarding the effects that restrictions on insurance legislation have on the decisions of physicians to certify sickness absence. It takes considerable time for changes in certification practices to be established and implemented not only among physicians, but also among patients who request sickness certificates.

The concept of disease in the sickness insurance system in Sweden

Many of the sickness insurance benefits which the Swedish social insurance system provides are linked to the individuals' earning income (Johansson & Jons-son, 1998), guaranteeing, to some extent, the possibility of maintaining the same standard of living. The development of sickness benefits during the 1990s in case of, for instance, long-term sickness absence are marked by a clear legislative ambition to exclude the possibility of individuals to receive benefits from sickness insurance for reasons other than "strictly" medical ones (Stendahl, 2003).

The National Insurance Act (AFL) is one of the important corner-stones in the Swedish welfare systems, and it stipulates the framework in which the social insurance authority must act. As a set of regulations, the Swedish sickness insurance has developed over many years and has been subjected to several changes during the years. This framework might at first sight be considered precise and detailed, but at a deeper investigation it has been found that it allows fairly broad adjustment of the norms at the level of supervision and application (Stendahl, 2003).

The term disease is not defined in the AFL. The statements made by the Social Insurance Committee (SOU, 1944) in the 1940s are still used as a guideline. That committee did not intend for its statements to be bound by court of law. Instead, they indicated that it was necessary to define the concept of disease through everyday language and by what medical science regard to be disease. The intent of that early Committee was that subsequent changes to policy would be easier to adapt to ongoing scientific knowledge in medicine, without changing an entire set of rules (SOU, 1995). Due to the increase of scientific knowledge and the medical achievements that occurred during the 1980s and 1990s, a growing number of physical complaints became accepted as medical conditions that entailed entitlement to sickness benefits (Nachemson & Jons-son, 2000).

In 1997, the term *disease* was further clarified in the official guidelines (SOU, 1997), and this time the focus was on the definition of work capacity as it related to the criterion of disease. The main imperative was that work capacity should be assessed on strictly “medical” grounds, which implies a narrower criterion than indicated in 1995. The change related to this criterion was described as a clarification, and the change related to the criterion of work capacity was explained as a better implementation of the existing rules. Today, entitlement to sickness absence benefits according to the National Insurance Act require a clear connection between the disease/illness/injury and the reduction in work capacity, although the concept of work capacity is not defined (National Insurance Act, (AFL, 3:7)).

Researchers (Timpka *et al.*, 1994; Cassis *et al.*, 1996) have underlined the gap between the importance of medical diagnosis in the health care services on the one hand and the social insurance authority on the other. The latter requires a precise evaluation of the incapacity to work, despite the inherent uncertainty of such assessments. Problems with sickness certification, including the assessment of the work capacity of the patients, often lead to more work for the physicians. An on-going discussion among physicians concerns the problems they experienced in connection with providing sickness certificates, the consequences of this practice, and different ways of dealing with the dilemma of trying to help each patient as much as possible while allocating limited resources (Hussey *et al.*, 2004; Arrelov *et al.*, 2005). Additionally, in Sweden and several other industrial nations, county councils or corresponding organisations (with which the health care services are affiliated) and sickness insurance systems have been developed and implemented their various programmes and policies in an uncoordinated manner.

Sickness certificates – is the content and form of the text of importance?

The sickness certificate is a major instrument for establishing contact and conveying information between the health care services and the social insurance authority. Obtaining a sickness certificate has a substantial impact on the life and work situation of the patient, as well as impacting on the economic costs of the society. In most Western countries, the responsibilities of physicians include assessing the ability of their patients to work and issuing sickness certificates when this capacity is reduced. The way these tasks are performed can strongly influence the situation of the patients and also affect the financial burden on society. In Sweden, the duties of physicians in conjunction with sick leave are regulated by the Health Service Act and by the National Board of Health and Welfare (Alexanderson & Norlund, 2004). The physician has two important

roles in their work with sickness certification. The first role is as the patient's physician and entails creating confidence and trust, understanding the patient, and being able to help in a meaningful way. The second role is that of a medical expert, in other words to objectively certify the patient's medical situation for other actors, such as the social insurance authority (Timpka *et al.*, 1994; Bredkjaer-Rask *et al.*, 1996; Maeland *et al.*, 2002). Accordingly, physicians have a gate-keeping role in ensuring that sickness certificates are issued only to those who have medically acceptable reasons for being absent from work (Sawney, 2002). A sickness certificate provided by a physician is to include information on the patient's occupation, his or her diagnosis, the degree of severity, and duration of reduced functional capacity, the prognosis for regain of functional capacity, and activities or rehabilitation measures that are recommended during the period of absence (Söderberg & Alexanderson, 2003; Alexanderson & Norlund, 2004).

A study (Hussey *et al.*, 2004) on sickness certification practices of physicians indicated that any changes to certification practices must take into account the range and complexity of factors that influence the physician's decision making regarding sickness status. Such changes will otherwise not result in improvement. While certifying physicians do not make the final decisions about awarding sickness benefits, it is recognised that the suggestions they include on the sickness certificates have a substantial impact on the judgements that are made in the social insurance authority (Timpka *et al.*, 1994; Hensing *et al.*, 1997). It is the responsibility of case officers, affiliated to the social insurance authority, to assess the right to sickness benefits and measures needed to promote RTW for a sickness benefit recipient.

There are thus far only few studies (Bredkjaer-Rask *et al.*, 1996; Maeland *et al.*, 2002; Söderberg & Alexanderson, 2005) that have addressed the question of the quality of the information on the sickness certificates as a basis for decisions on sickness insurance benefits. The authors revealed that the certificates, in general, gave little insight into the normal work tasks of the patients or whether any type of rehabilitation measures had been considered. The main focus was on stating the diagnosis and on prolongation of a sick leave period. Regardless of the intended purpose of information on patients that is provided by the health care organisation, it should be used not only to guarantee appropriate and safe care, but also to serve as adequate guidance for actions to be taken (Thornquist, 2001). Furthermore, failure to address important workplace and other psychosocial variables can actually contribute to prolongation of sick leave periods (Plomp, 1999; Salazar & Graham, 1999). The lack of information in sickness certificates might also result in fewer personal meetings between the actors due to passive prolongations of sick leave periods, and a lack of possibilities for the individual to participate in face by face meetings.

The “work-line strategy”

During the last decade, many welfare states have become more active in developing strategies to promote RTW among sickness benefit recipients (Lierop & Nijhuis, 2000; Buys & Rennie, 2001). In Sweden and Norway, this reform has been called the “work-line strategy,” incorporated into the social insurance system at the beginning of the 1990s. The objective of RTW measures is to enable for individuals on sick leave to rejoin the labour market. Examples of these measures are work training and education. The principles of the “work-line strategy” were a clear manifestation of a shift from policies focusing on benefits and compensation towards more work-oriented policies, emphasised through administrative measures, coordination and early rehabilitation. As a result of this strategy, the tasks of the case officers in the social insurance administration have been broadened to include not only making decisions about sickness benefits, but also to assess the need for further measures, to facilitate RTW among sick-listed and if necessary, to coordinate such measures (Lindqvist, 2000).

According to Lindqvist and Marklund (1995) this strategy symbolises a new relationship between work and welfare. From the viewpoint of the state, this strategy means a decrease in sickness insurance benefits and disability pensions. However, an important precondition for success is that a multitude of different actors work in the same direction according to this strategy.

The employer’s responsibility for work environment in Sweden is regulated by both the Work Environment Act (AML) and the National Insurance Act (AFL). Under the latter, the employers bear the primary responsibility for detecting and investigating the need for RTW measures, ensuring that measures are implemented, and financing them. Furthermore, according to the National Insurance Act (AFL) the employer must carry out rehabilitation investigations in the event of sickness absence lasting for more than eight weeks. The aim of such investigation is to clearly identify RTW measures and other types of support in order to facilitate RTW, and to forward these to the social insurance authority within eight weeks from the beginning of a sick leave spell. Rehabilitation investigations appear, however, to be not very common in practice. It is evident from several studies (Hensing *et al.*, 1997; Marklund, 2000) that the responsibility for rehabilitation investigations is poorly understood and implemented.

According to a case study (Försäkringskassan, 2005) it is probable that a number of rehabilitation investigations are carried out without either a purpose or substance. They seem to be performed simply because they must be carried out. The results showed that the employers’ proposals in regard to training, part-time sick leave and RTW measures at another workplace are insubstantial and often concentrated on the individual’s incapacity.

There are probably other problems related to the rehabilitation investigations, such as a lack of focus on the actual need of changes to the individual's workplace or on measures facilitating the RTW process. Considering the amount of attention devoted to the potential of RTW measures, there is a notable lack of studies focusing on *how* to identify the sick-listed individuals who might profit from modified work conditions.

Management routines of the actors involved

Many documents and referrals in the health care system, and many investigations in the social insurance authority, are issued without connection to the workplace of the individual. In the assessments of documents (e.g. sickness certificates and rehabilitation investigations) it is expected that the use of issued information will decrease the risk of prejudicial treatment of the involved individuals. Furthermore, current sick leave reduction strategies rely on the assumption that the medical information in sickness certificates provides the case officers with sufficient information to identify sick-listed individuals for relevant and adequate RTW measures. Studies (Fleten, 2006) have however found, that despite the resources spent on sickness certificates, there is little information about the effectiveness of certificates as guidelines, for this assumption have not been tested according to daily practice. Thus far, scientific knowledge is lacking on how decisions regarding entitlement to sickness benefits and RTW measures are affected by the scarce information provided by health and medical services and employers.

Along with the "work-line strategy" the quality of the management routines of the involved actors (i.e. how and when they perform their assigned tasks) is an extremely important aspect affecting both the right to benefits and initiation of appropriate rehabilitation measures that can facilitate RTW. A period of sick leave of the individual include contact with different actors and can be summarised by the following characteristics: an authority driven structure, mostly based on written documents providing information that often is fragmentary, centred on diagnosis aimed at stating the incapacity of the individual, and with very little concern for matching the patient's workplace requirements to the RTW process.

The work capacity of the individual

The RTW process is intended to include all those activities that are designed to analyse the starting point of the sick-listed individual and to use that assessment as a basis for developing an individual plan for reintegration in the labour market (SOU, 2000). However, before this planning condition is reached, the paper-

driven process, consisting of sickness certificates and rehabilitation investigations, appears to proceed independently of medical rehabilitation. Previous studies (SBU, 2004) have documented problems in assessing the optimal length and degree of sickness certification including a lack of valid, reliable instruments for assessing work capacity. Furthermore, the different views on disease and work capacity among the actors involved (e.g. the health and medical service and the social insurance authority) may give rise to different criteria being used in the assessment of the work capacity of the individuals. Different conceptions on disease and work capacity may also lead to differences in individuals' roles and degrees of participation during the RTW process (SOU, 2000; Krause *et al.*, 2001).

Notwithstanding, during the last two years there has been a change in the practices of case officers affiliated to the social insurance authority in Sweden. Case officers (Försäkringskassan, 2005) have stated that they more often withdraw an application for sickness benefits nowadays due to the fact that patients are deemed not to have medically acceptable reasons for being absent from work.

The demands of the case officers have increased in recent years, since it was decided that entitlement to sickness benefits should be made strictly on medical grounds and that sickness certificates should be thoroughly assessed before deciding to grant compensation. Surprisingly little research has examined the experiences of case officers in the domain of social insurance (Hensing *et al.*, 1997; Söderberg & Alexanderson 2005). The conclusion that can be drawn is that we know all too little about the practices connected with the assessment of work capacity and decisions regarding sickness benefits.

The coordination meeting

Coordination of RTW measures in the social insurance authority involve new methods of working, and according to present law and regulations, AFL 3:8, a coordination meeting with the individual is to take place 90 days after the rehabilitation investigation have been performed. At this meeting, several participants are expected to take part, for instance, the certifying physician, the employer, the occupational health service, the employment office in case of unemployment, along with the individual and representatives from the social insurance authority. The object of such meetings is to enhance the contact between the actors involved and to clarify any issues regarding RTW status of the individual.

According to a study (Försäkringskassan, 2005) investigating coordination meetings the majority of case officers thought that such meetings were a good way of promoting return to work. In their view the coordination meeting should lead to a faster return to the previous workplace, either on full or part-

time status, or the employee's seeking new work. The same findings were elucidated in a study (Söderberg *et al.*, 2006) on sickness certification practice of general practitioners (GPs). Asked for their preferences regarding future contacts with the social insurance authority, the GPs suggested more coordinating meetings and a contact person at the social insurance office that they could quickly reach when necessary.

On the whole, identification of factors related to RTW by an individual is a multifaceted task, and thus handling the contact with the individual is an essential challenge for the actors involved in this process.

Implications for future research

There are no simple and clear-cut guidelines for successful RTW. Individuals differ widely with respect to personal needs, desire to be helped and supported, as well as regarding at what phase of the rehabilitation process they want to interact with professionals. This indicates that RTW efforts and interactions between different actors and the sick-listed individual are complicated matters that require substantial skill and knowledge on the part of the actors involved. The present system seems to place little emphasis on establishing the prerequisites for such professional interactions as a way of supporting individuals. The scientific knowledge base in this area needs to be extended to provide new perspectives on management routines and interaction between the people involved – both the professionals and the patients/clients.

In Sweden in particular, there is a chance that the health care services and the social insurance authority might take over the responsibility for the RTW process, and that their plans for action might be contradictory to the development of workplace related measures. For a previously sick-listed individual to rejoin the workplace may therefore grow more difficult if sick leaves are considered too lengthy, and there is a risk that a disability pension may come to replace a return to work. Furthermore, individuals on long term sickness absence may experience the risk of being increasingly distanced from the sphere of work life. This in turn might entail loss of social network and loss of control over one's own life. The employer's prejudice against long leaves of absence due to sickness might place an individual in an awkward position of sometimes trying to appear to be more ill and disabled than is really the case.

It is essential to study the cooperation between the individual, the physician, and the case officer, to develop more understanding on the RTW cycle, from both process and system perspectives. Further research is also needed to discern the importance of gender in RTW-promoting measures, because there is still a lack of knowledge about how gender is “constructed” in the interaction between the actors and the individual, especially in light of the reported gender-

related differences in sickness absence (Kilbom *et al.*, 2001; Wamala & Lynch, 2002). Future studies should also try to better explain the gate-keeping and RTW-promoting functions of case officers. Research on this entire course of activities is scarce; neither does it appear that the theoretical or methodological aspects of this process have been investigated from the perspective of the daily practices of the actors (e.g. physicians and case officers) involved.

References

- Aarts LJM, Burkhauser RV, *et al.* (1996). *Curing the Dutch Disease: an International Perspective on Disability Policy Reform*. Ashgate, Avebury, Aldershot UK and Brookfield.
- Alexanderson K, Norlund A (2004). Sickness absence - causes, consequences, and physicians' sickness certification practice. A systematic review by the Swedish Council of Technology Assessment in Health Care. *Scandinavian Journal of Public Health*, Vol. 32 (Suppl. 63): 12-258.
- Arrelov B, Borgquist L, *et al.* (2005). Influence of local structural factors on physicians' sick-listing practice: a population-based study. *European Journal of Public Health*, Vol. 15 (5): 470-474.
- Brage S, Bentsen B, *et al.* (1996). ICPC as a standard classification in Norway. *Family Practice*, Vol. 13 (4): 391-396.
- Bredkjaer-Rask S, Hvas AC, *et al.* (1996). Er Den Udvidede Attest LAE 355 Anvendelig ved Langvarige Sygedagpengesager? (Is the extended certificate usable in long-term sickness absence). *Ugeskr Laeger*, Vol. 158 (5): 603-606. [In Danish]
- Buyt NJ, Rennie J (2001). Developing relationships between vocational rehabilitation agencies and employers. *Rehabilitation Counseling Bulletin*, Vol. 44 (2): 95-103.
- Cassis I, Dupriez K, *et al.* (1996). Quality of work incapacity assessment in the Swiss disability insurance system. *International Journal for Quality in Health Care*, Vol. 8 (6): 567-575.
- Englund L, Tibblin G, *et al.* (2000). Effects on physicians' sick-listing practice of an administrative reform narrowing sick-listing benefits. *Scandinavian Journal of Primary Health Care*, Vol. 18 (4): 215-9.
- Fleten N (2006). *The sick-listed - an under-recognised resource in handling sickness absence*. Dissertation thesis. Tromsø: Faculty of Medicine.
- Friesen M, Yassi A, *et al.* (2001). Return to work: The importance of human interactions and organization structures. *Work*, Vol. 17 (1): 11-22.
- Försäkringskassan (2005). *Avstämningsmöte. Försäkringskassans metodundersökning 2004 (Statusmeeting. Investigation on methods 2004)*. Stockholm: Försäkringskassan. [In Swedish]

- Försäkringskassan (2005). *Bedömning av rätten till sjukpenning (Assessments on the right to sickness benefits)*. Linköping: Försäkringskassan, Landstinget i Östergötland och Linköpings Universitet. [In Swedish]
- Försäkringskassan (2005). *Rehabilitationsutredning - plan för återgång i arbete (Rehabilitation investigation - plan for RTW)*. Stockholm: Försäkringskassan. [In Swedish]
- Getz L, Westin S (1994). Behandler og sakkyndig - mellom barken og veden? (Doctors and the disability pension). *Tidsskr Nor Lægeforen*, Vol. 114 (12): 1435-1440. [In Norwegian]
- Goine H (2006). *Sickness absence - aspects of Measurement, Impact of the Labour Market and Effects of Interventions*. Östersund: Department of Health Sciences, Mittuniversitetet.
- Hensing G, Alexanderson K, *et al.* (1997). Dilemmas in the daily work of social insurance officers. *Scandinavian Journal of Social Welfare*, Vol. 6: 301-309.
- Himmel W, Sandholzer H, *et al.* (1995). Sickness certification in general practice. *European Journal of General Practice*, Vol. 1: 161-166.
- Hussey S, Hoddinott P, *et al.* (2004). Sickness certification system in the United Kingdom: quality study of views of general practitioners in Scotland. *British Medical Journal*, Vol. 328 (7431): 88-91.
- Johansson S, Jonsson M (1998). Vocabularies and citizenship and gender: Sweden. *Critical Social Policy. A Journal of Theory and Practice in Social Welfare*, Vol. 18: 39-415.
- Jönsson B, Arvidsson G, *et al.* (2004). *Hälsa, vård och tillväxt. Välfärdspolitiska rådets rapport 2004 (Health, Care and Development)*. Stockholm: SNS Förlag. [In Swedish]
- Kilbom Å, Messing K, *et al.* (2001). *Women's health at work*. Stockholm: National Institute of Working life.
- Krause N, Dasinger L, *et al.* (2001). Psychosocial job factors and return-to-work after compensated low back injury: A disability phase-specific analysis. *American Journal of Industrial Medicine*, Vol. 40 (4): 374-392.
- Lierop B, Nijhuis F (2000). Assessment, education and placement: an integrated approach to vocational rehabilitation. *International Journal of Rehabilitation Research*, Vol. 23: 261-269.
- Lindqvist R (2000). *Att sätta gränser. Organisationer och reformer i arbetsrehabilitering (Creating borders. Organisations and reforms in vocational rehabilitation)*. Umeå, Boréa Bokförlag. [In Swedish]
- Lindqvist R, Marklund S (1995). Forced to work and liberated from work - a historical perspective on work and welfare in Sweden. *Scandinavian Journal of Social Welfare*, Vol. 4 (4): 224-237.
- Maeland JG, Ringdal PR, *et al.* (2002). Kvalitetsvurdering av Sykmelding II. (Assessing the quality of medical certificates in long-term sickness absence). *Tidsskr Nor Lægeforen*, Vol. 122: 161-4. [In Norwegian]

- Marklund, S., Ed. (2000). *Arbetsliv och hälsa 2000 (Worklife and Health in Sweden 2000)*. Solna, Arbetslivsinstitutet, Arbetsrsyddstyrelsen. [In Swedish]
- Nachemson A, Jonsson EE (2000). *Neck and Back Pain. The scientific evidence of Causes, Diagnoses, and Treatment*. Philadelphia: Lippincott Williams & Wilkins.
- Overbye E (2005). Dilemmas in Disability Activation and How Scandinavians Try to Live with Them. Social Security and Participation in Economic and Social Life. In: *Welfare to Work in Practice*. Saunders P, editor. Hants and Burlington: Ashgate Publishing.
- Plomp HN (1999). Evaluation of doctor-worker encounters in occupational health: an explanatory study. *Occupational Medicine*, Vol. 49 (3): 183-188.
- Salazar M, Graham K (1999). Evaluation of case management program: Summary and integration of findings. *American Association of Occupational Health Nurses Journal*, Vol. 47: 416 -423.
- Sawney P (2002). Current issues in fitness for work certification. *British Journal of General Practice*, Vol. 52: 217-222.
- SBU (2004). *Prescribed Sick Leave - Causes, Consequences and Practices. A systematic review*. Stockholm: The Swedish Council of Technology Assessment in Health Care. Pp: 8-9.
- Shiels C, Gabbay M, et al. (2004). Patient factor associated with duration of certified sickness absence and transition to long-term incapacity. *British Journal of General Practice*, Vol. 54: 86-91.
- Socialdepartementet (2003). *Den svenska sjukan II - regelverk och försäkringsmedicinska bedömningar i åtta länder (The Swedish disease - sickness absence in eight countries)*. Stockholm: Studier av offentlig ekonomi, RFV och Socialdepartementet. [In Swedish]
- SOU (1944). *Socialvårdskommitténs betänkande VII: Utredning och förslag angående Lag om Allmän sjukförsäkring (Inquiry and consideration concerning the National Insurance Act, AFL)*. Stockholm: Socialdepartementet. [In Swedish]
- SOU (1995). *Försäkringskydd vid sjukdom - ett delbetänkande om rätten till ersättning och beräkning av inkomstunderlag under sjukpenningtid (Insurance cover at disease - a consideration on the entitlement to compensation and calculation of income qualifying during sickness allowance period)*. Stockholm: Socialdepartementet. [In Swedish]
- SOU (1997). *Obälsförsäkringen - Trygghet och aktivitet (Disease insurance - Security and activity)*. Stockholm: Socialdepartementet. [In Swedish]
- SOU (2000). *Rehabilitering till arbete - en reform med individen i centrum (Rehabilitation for work - a reform centred on the individual)*. Stockholm, Socialdepartementet. [In Swedish]
- Stendahl S (2003). *Communicating Justice Providing Legitimacy - The legal practices of Swedish administrative courts in cases regarding sickness cash benefits*. Göteborg: Department of Law, Göteborgs Universitet.

- Söderberg E (2005). *Sickness Benefits and Measures Promoting Return to Work - Perspectives of Different Actors*. Dissertation thesis. Linköping: Linköping University.
- Söderberg E, Alexanderson K (2003). Sickness certification practices of physicians - a review of the literature. *Scandinavian Journal of Public Health*, Vol. 31 (6): 460-474.
- Söderberg E, Alexanderson K (2005). Gate-keepers in sickness insurance - a systematic review of the literature on practices of social insurance officers. *Health and Social Care in the Community*, Vol. 13 (3): 211-223.
- Söderberg E, Alexanderson K (2005). Sickness certificates as a basis for decisions regarding entitlement to sickness insurance benefits. *Scandinavian Journal of Public Health*, Vol. 33 (4): 314-320.
- Söderberg E, Hagberg J, *et al.* (2006). Vårdcentralsläkares upplevelser av kontakter med försäkringskassan samt önskemål om hur dessa kontakter kan förbättras (GPs' sickness certification practice - problems experienced in contact with the social insurance office and preferences for future such contacts). In: *Läkarens arbete med sjukskrivning*. Löfgren A, editor. Stockholm: Sektionen för personskadeprevention, Institutionen för klinisk neurovetenskap, Karolinska Institutet. [In Swedish]
- Thornquist E (2001). *Kommunikation i kliniken (Communication in the clinic)*. Lund: Studentlitteratur. [In Swedish]
- Timpka T, Hensing G, *et al.* (1994). Dilemmas in Sickness Certification among Swedish Physicians. *European Journal of Public Health*, Vol. 5: 215-219.
- Wadell G, Aylward M, *et al.* (2002). *Back pain, incapacity for work and social security benefits. An international literature review and analysis*. London: The Royal Society of Medicine Press.
- Wamala S, Lynch J (2002). *Gender and Social Inequities in Health - a public health issue*. Lund: Studentlitteratur.