

From the Department of Clinical Sciences, Family Medicine Stockholm, Karolinska
Institutet, Stockholm

General Practitioners' Decision-Making on Drug Treatment of Hypercholesterolaemia

Lars Backlund



Stockholm 2003

All previously published papers were reproduced with permission from the publisher.

Published and printed by Karolinska University Press

Box 200, SE-171 77 Stockholm, Sweden

© Lars Backlund, 2003

ISBN 91-7349-672-3

To Lena, Tobias, David, Fanny and Filippa

Abstract

Drug treatment of elevated cholesterol values may be indicated for individuals at high risk for cardiovascular disease and as a complement to lifestyle advice. Guidelines recommend a numerical risk calculation as the basis for selecting the individuals at highest risk among those who are free from previous cardiovascular disease (primary prevention). Guidelines and tools for risk estimation are often not available or not used, which leads to a risk for over- and undertreatment.

The purpose was to examine General Practitioners' (GPs) decisions on drug treatment for patients with hypercholesterolaemia. Written patient cases were presented either in a paper format (Studies I and II) or on a computer screen (Studies III-V). In Study I, 38 GPs rated their inclination for drug treatment for 40 case descriptions. Each doctor's strategy was defined as the set of statistical regression weights for the different variables describing the patient. The strategies varied widely between doctors. The most important variable was previous coronary heart disease, followed by the degree of cholesterol elevation. The majority of GPs used two or three of the eight variables in their judgements. In Study II the ability of GPs and medical students to make risk estimates without using risk assessment tools was examined. Both groups underestimated risk, especially for high-risk patients.

In studies III-V, think-aloud technique was used to examine the thought processes leading to the decision to prescribe or not. The GPs were instructed to say aloud all their thoughts about the information presented on successive computer screens. The audio-taped and transcribed protocols were coded for whether the statements favoured or disfavoured drug prescription (directionality of decision). They were also coded regarding both the information about the patient to which they referred and their cognitive content. In Study III, half of the participants not only talked aloud but also rated their inclination toward drug prescription at each of the successive screens with new information. The ratings and verbal protocols reflected the change in directionality in similar ways, which was interpreted as supporting their validity. Verbal protocols were at least as sensitive as ratings in reflecting the change in decision directionality over time.

In studies IV and V the think-aloud data were analysed regarding how different kinds of information about the cases were evaluated and used in the decisions. Cholesterol level was most important and different cut-off levels were used by different GPs. Lifestyle-related factors seemed to be evaluated from different perspectives. A patient's smoking could be regarded as increasing the risk and thereby favouring drug prescription by some GPs, and as a possibility for lifestyle change and as an argument for refraining from treatment by other GPs. After the six cases, the GPs were asked to describe in their own words their thinking concerning treatment of hypercholesterolaemia. These protocols were coded for cause-effect relations to map the doctors' knowledge and opinions. The GPs made several departures from a strict application of guidelines in the individual cases, even when they had expressed the relevant knowledge contained in the guidelines. Their arguments often concerned lifestyle factors.

The think-aloud technique with the coding scheme developed for the present studies seemed to be a valuable complement to statistical approaches to judgements and decisions (Studies I-II). One possible application area is studies on differences in decision processes of experienced doctors and medical students, the results of which can be used in medical training. Another application area involves the usability and efficiency of different decision support systems integrated in the computerised medical record.

Key words: general practice, guidelines, decision-making, judgements, Clinical Judgement Analysis, case vignettes, hypercholesterolaemia, think-aloud.

List of Publications

This thesis is based on the following papers, which will be referred to by their Roman numerals.

- I. Backlund L, Danielsson B, Bring J, Strender L-E.
Factors influencing GPs' decisions on the treatment of hypercholesterolaemic patients.
Scandinavian Journal of Primary Health Care 2000; 18: 87-93.
- II. Backlund L, Bring J, Strender L-E. GPs and Students underestimate coronary risk in hypercholesterolaemic patients.
Primary Health Care Research and Development (In press).
- III. Backlund L, Skånér Y, Montgomery H, Bring J, Strender L-E.
Doctors' decision processes in a drug prescription task: The validity of rating scales and think-aloud reports.
Organizational Behavior and Human Decision Processes 2003; 91:108-117.
- IV. Backlund L, Skånér Y, Montgomery H, Bring J, Strender L-E. GPs' decisions on drug treatment for patients with high cholesterol values: A think-aloud study. Submitted manuscript
- V. Backlund L, Skånér Y, Montgomery H, Bring J, Strender L-E.
The role of guidelines and the patient's life-style in GPs' management of hypercholesterolaemia. Submitted manuscript.

Contents

Abstract.....	v
List of Publications	vi
Contents	7
List of Abbreviations.....	9
Introduction	10
The Field of Study and Some Notes on Terminology	10
Risk Factors and Guidelines	11
Risk Factors	11
Formulas and Charts for Risk Estimation.....	12
Guidelines	13
Guidelines on Lipid Management.....	15
GPs' Use of Guidelines on Lipid Management.....	18
Doctors' Judgements and Decisions.....	18
Risk Estimates	19
Weighting of Evidence, Clinical Judgement Analysis.....	20
Decision-Making as Cognitive Processes.....	23
Think-Along Technique Applied to Medical Decision-Making	24
Process Theories on Decision Making.....	25
Aims.....	27
Methods	28
Patient Cases/Case Vignettes	28
Participants	28
Design, Procedure and Data Analyses.....	29
Study I.....	29
Study II.....	30
Studies III-V	32
Results	36
Study I.....	36
Response Rate.....	36
Mean Ratings	36
Regression Weights	36
Relation Between Actual and Stated Strategies	36
Comparison with Guidelines.....	37
Study II	38
Response Rate.....	38
Risk Estimates for Different Cases	38

Isolated Hypercholesterolaemia	38
Accuracy	38
Treatment Decisions	40
Study III	40
Interjudge Reliability of Coding	40
Comparing Ratings and Think-Aloud Protocols (Validity)	40
Effects of Adding a Rating Scale	43
Study IV	43
Reliability of Coding	43
Information Categories	43
Treatment Decisions	44
Importance of Information	44
Patterns of Importance for “Yes” and “No”- Decisions	44
Clustering of Participants	45
Disagreement	46
Use of Rules	46
Study V	46
Contents of the Free-Report Protocols	46
Evidence of Knowledge Related to Guidelines	47
Decisions and Arguments	47
Discussion	49
Information Affecting the Decisions	49
Evaluation and Integration of Information	50
Decisions and Risk Estimates Compared with Guidelines	52
Methodological Considerations	53
Participating Doctors	53
Validity of Case Vignettes	53
Evaluations of the Decisions	55
Conclusions	57
Forthcoming Research	58
Acknowledgements	59
References	61
Summary in Swedish – Sammanfattning på svenska	71
Papers	73

List of Abbreviations

CHD	Coronary heart disease
CJA	Clinical judgement analysis
CVD	Cardiovascular disease
GP	General practitioner
HDL	High density lipoproteins
LDL	Low density lipoproteins
NNT	Number needed to treat
PAD	Peripheral artery disease
TIA	Transient ischaemic attack

Introduction

The Field of Study and Some Notes on Terminology

This thesis is about medical decision-making, which involves how medical professionals make decisions about treatment, diagnosis, the ordering of tests, referrals and other medical tasks. A decision can be conceived of as a choice between different options for action or for preference. A decision is typically preceded by a judgement of the decision alternatives, although the distinction between decision and judgement is not always clear.

The medical area examined in this thesis is treatment of elevated cholesterol values. The terms high cholesterol or hypercholesterolaemia have sometimes been used instead of the more comprehensive and correct terms elevated blood lipids and hyperlipidaemia. In the language of ordinary clinical practice and according to how the laboratory investigations are labelled, blood lipids is a general term for total cholesterol (which here and in clinical language is often replaced by the term cholesterol), triglycerides, low density lipoproteins (LDL), and high density lipoproteins (HDL).

Total cholesterol and LDL are both generally regarded as belonging to the major risk factors for arteriosclerotic disease, and they have been used with different relative emphasis on one or the other in studies on treatment effects and in guidelines on lipid management. Increased levels of triglycerides and decreased levels of HDL have been regarded as markers for additional risk for future arteriosclerotic disease, but they have not been used as major criteria in guidelines or for the selection of patients in studies.

Most intervention studies and guidelines have used coronary heart disease (CHD) as the outcome, i.e. the disease to be prevented. CHD ordinarily includes myocardial infarction, angina pectoris, silent myocardial ischaemia, and sudden death. The more general term cardiovascular disease (CVD) is sometimes used in the thesis and it includes stroke and transient ischaemic attacks (TIA) and peripheral artery disease (PAD), in particular arteriosclerosis in the lower extremities (intermittent claudication).

The focus of the thesis is on the decisions made by general practitioners (GPs) as to whether to initiate drug treatment or not depending on the characteristics of different cases. The judgements and decisions were made in response to written patient cases that were generally based on authentic patients from primary health care. The most important information about the patient cases was presented either in a paper format or on a computer screen. Patient vignettes, case vignettes and

written case descriptions are some of the terms used for this technique in the literature and in the present text.

Medical decisions are the concern of all categories of health care personnel. For the sake of simplicity, however, and because the majority of studies on medical decision-making have focused on doctors, the words physician and doctor have generally been used to identify the medical decision-maker. A decision on a group level is not typical for the work of GPs and has not been a concern in the studies reported here.

Risk Factors and Guidelines

Risk Factors

CHD is the leading cause of premature death and morbidity in Western societies (Murray 1997). The probability for an individual to develop a coronary event is determined by several risk factors that interact to produce a total risk. It is therefore important for health care personnel to be able to evaluate an individual's risk level and to acquire strategies for reducing risk. Some of the risk factors are not modifiable, such as age, sex and family history. The most well recognised modifiable risk factors are hypertension, diabetes, hyperlipidaemia, smoking, overweight and physical inactivity.

Although it has often been stated that half of the patients who develop CHD are free from known risk factors, a recent survey of studies (Greenland 2003) showed that almost every patient who developed CHD had at least one of the following risk factors: diabetes, cigarette smoking, hypertension or cholesterol elevation. The prevalence of modifiable risk factors among patients with CHD was demonstrated to be high according to a 1995-1996 survey of nine European countries concerning hospitalised patients with CHD (EUROASPIRE Study Group 1997). For example, 19% were smokers and 86% had elevated cholesterol values. In a follow-up in 1999-2000 (EUROASPIRE I and II Group 2001), the proportion of smokers remained almost unchanged, while the prevalence of obesity (body mass index ≥ 30 kg/m²) increased from 25% to 33%. The proportion of patients with high cholesterol values had decreased substantially. However, among these patients with CHD, 59% had not attained adequate lipid control. Among a primary health care population in the Netherlands derived from both high-risk and low- to normal-risk categories, cholesterol levels increased between 1977 and 1995 (Bakx 2000), which seemed to be caused in part by increased weight.

On a population level, The WHO MONICA project has described the trends in coronary disease and coronary risk factors. The risk factors seemed to be

decreasing in most populations between 1979 and 1996 (Evans 2001), but worrying findings were tendencies for an increase in smoking among women and an increase in weight, especially in men. The mean cholesterol level showed a small downward trend. The trend regarding the rate of CHD seems to vary in different populations, with a decrease over time in Sweden (Truelsen 2003).

Increasing activity on the part of health care personnel to reduce CHD risk factors among both the high-risk individuals and the population as a whole should have great potential to reduce CHD. The construction and implementation of treatment guidelines for preventing CHD, with lipid control as one of the targets, should be one of the most important activities for attaining this goal. All the recent guidelines on management of coronary risk factors include the concept of numerical risk estimation, and this concept will be discussed first.

Formulas and Charts for Risk Estimation

For patients without previous CVD, i.e. the focus of primary prevention, the guidelines on treatment of hyperlipidaemia generally recommend a calculation of the absolute risk for CHD. With a risk above a certain cut-off value, pharmacological treatment is generally justified or "should be considered", if lifestyle intervention has not succeeded in normalising the cholesterol value. Almost all the equations for risk calculation are based on the so-called Framingham equation from the Framingham Heart Study (Dawber 1951). Quite recently, however, prediction algorithms have been published that are based on databases from several European populations (Hense 2003, Third Joint Task Force of European and Other Societies on Cardiovascular Disease Prevention in Clinical Practice 2003). The Framingham Heart Study started in 1948 and has since been subjected to frequent follow-ups. It was the first multifactorial approach to cardiovascular disease and it introduced the concept of risk factors. Since the 1970s prediction equations have been modelled on the basis of the outcomes from the study. The most frequently used algorithms are the equations by Anderson and co-workers (Anderson 1990, 1991) although other types of equations have been formulated. For a comparison between different multivariate methods, see Knuiman (1997). The equations allow for the calculation of end-points at different times and for different disease endpoints. Eight different risk factors are used: age, sex, systolic or diastolic blood pressure, total cholesterol and HDL, presence or absence of left ventricular hypertrophy, diabetes mellitus, and cigarette smoking. The model that is used is a parametric regression model for risk estimation.

To avoid the need for formal risk calculation, several risk charts or risk tables have been developed. The charts are simplified in comparison with the

Framingham equation upon which they are based. They differ from one another with regard to which risk factors are included and where the cut-off value for numerical risk is set. A 10- year risks for CHD of 20% or higher is suggested in the Swedish guidelines (Läkemedelsverket 1999), the joint European guidelines (Wood 1998) and the US guidelines from 2001 (Smith 2001). As a comparison, the British Department of Health recommends the use of Sheffield tables with a risk of 30% over 10 years (Ramsay 1996), and the joint British Societies (British Cardiac Society 1998) recommend a 10- year CHD risk of 15%. As an illustration of a risk chart, Figure 1 shows how the 10-year risk for women can be calculated according to the 1998 European guidelines.

The usefulness of the risk charts depends on their correspondence with the Framingham equation. A validation of this kind against the Framingham equation for different charts developed in Britain, Canada, New Zealand and the joint European charts was undertaken (Jones 2001) using 691 patients in primary health care. The best correspondence with the Framingham equation was achieved by the joint British Guidelines (British Cardiac Society 1998) with the following risk factors: age, sex, smoking, systolic blood pressure, diabetes, total cholesterol/HDL ratio.

The criteria for selecting patients for pharmacological treatment should be based not only on the size of the medical effects for different patient groups, but also on health economic considerations. Drug treatment for secondary prevention is generally considered cost-effective (Johannesson 1997). In primary prevention, which may involve a large proportion of the population, the issue of cost-effectiveness comes into focus. The long-used 10-year risk of 20% seems to have been selected without explicit health-economic considerations. Johannesson (2001) carried out health-economic evaluations of the 4S study and demonstrated that the cut-off value for when treatment is cost-effective seems to vary with age and gender, and with the levels of other coronary risk factors, with better cost-benefit for higher risk. Prosser (2000) also showed a higher cost-effectiveness with a higher level of risk.

Guidelines

Guidelines have been defined as “systematically developed statements to assist practitioner and patient decisions about appropriate healthcare for specific clinical circumstances” (Field 1990). The purpose of guidelines is to improve practice, reduce its variability and improve health care outcomes. They should be based on scientific evidence and take into account health economy as well as the local

Guidelines on Lipid Management

Guidelines on lipid management have undergone repeated modifications during the past decade, both in Sweden and internationally, mainly due to the influence of the outcomes of large intervention studies. During this same period of time a number of studies have been published that have demonstrated the positive effects of statines (3-hydroxy-3 methylglutaryl coenzyme A reductase inhibitors) on CHD.

Four large studies on secondary prevention have evaluated the effects of statines on coronary events: the 4S study (Scandinavian Simvastatin Survival Study Group 1994), CARE (the Cholesterol and Recurrent Event Trial Investigators 1996), LIPID (Long-Time Intervention with Pravastatin in Ischemic Disease (LIPID) Study Group 1998) and HPS (Heart Protection Study Collaborative Group 2002). They were all randomised with large study groups (between approximately 4000 and 20 000 individuals) including both men and women, and the effects were studied for five to six years.

The relative reduction in LDL was 35% in the 4S study, 31% in CARE, 26% in LIPID and 21% in HPS. The corresponding relative reductions in myocardial infarctions were 34% in 4S, 24% in CARE and LIPID and 26% in HPS. The reduction in absolute risk was largest in 4S (8.6%) and varied from 3.0% to 3.6% in the other studies.

The two most important primary prevention studies with statines are WOSCOPS (the West of Scotland Coronary Prevention Study Group 1995), including only men, and AFCAPS/TexCAPS (the AFCAPS/TexCAPS Research Group 1998). The relative reduction in LDL was 18% in WOSCOPS and 25% in AFCAPS/TexCAPS. The corresponding relative risk reductions for coronary events were 31% and 40%, respectively, while the absolute risk reductions were lower than in the secondary prevention trials at 2.4% in WOSCOPS and 2.3% in AFCAPS/TexCAPS.

Due to the lower initial risk for patients in primary prevention, and correspondingly less reduction in absolute risk, the number of individuals who need to be treated (NNT) for five years in order to prevent one coronary event for one individual is higher than in secondary prevention. In the studies cited, the NNT in primary prevention was 43 and 45, respectively, and in secondary prevention it varied between 13 (4S) and 33 (CARE).

The first Swedish guidelines on cholesterol treatment were published in 1988 (Socialstyrelsens läkemedelsavdelning). The guidelines were later modified (Läkemedelsverket 1995) taking into account the 4S-study and recommending a

lower cut-off value for total cholesterol for initiating drug treatment (5.5 mmol/l). They also stressed the importance of previous CHD. Figure 2 shows a simplified summary of the message in these 1995 guidelines.

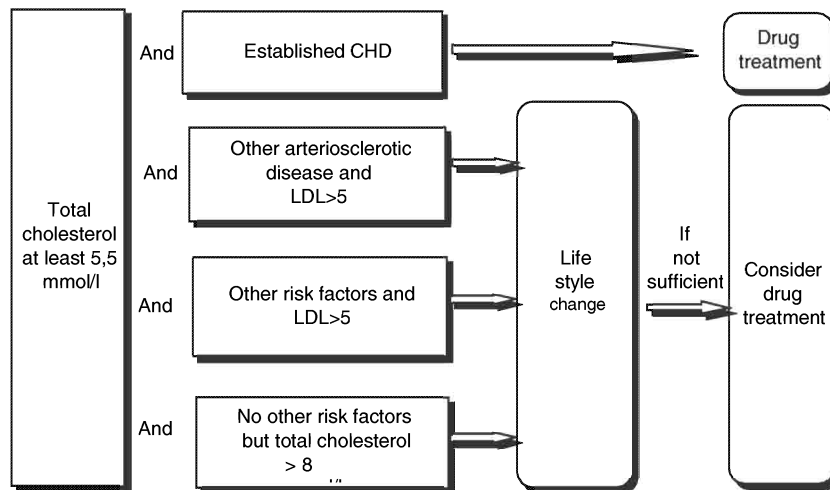


Figure 2. A simplified version of the 1995 Swedish guidelines on treatment of hyperlipidaemia.

The Swedish guidelines were changed a few years later (Läkemedelsverket 1999) as a modification of the joint European recommendations (Wood 1998). Some of the main characteristics of these Swedish guidelines are summarised in Figure 3.

Both the Swedish and European guidelines state that the goal values for treatment and the cut-off values for initiating treatment should be the same for all individuals: 5.0 mmol/l for total cholesterol and 3.0 mmol/l for LDL. The primary target is a patient with CHD or other arteriosclerotic manifestations. For individuals without these diseases a quantitative risk estimate is recommended based on the Framingham study (e.g. Anderson 1990, 1991). Information about the individual's age, total cholesterol, blood pressure, and smoking status is entered on a chart and the resulting 10-year risk for a coronary event is then determined from the chart. Separate charts are used for men and women. If the 10- year risk exceeds 20%, pharmacological treatment is generally advised.

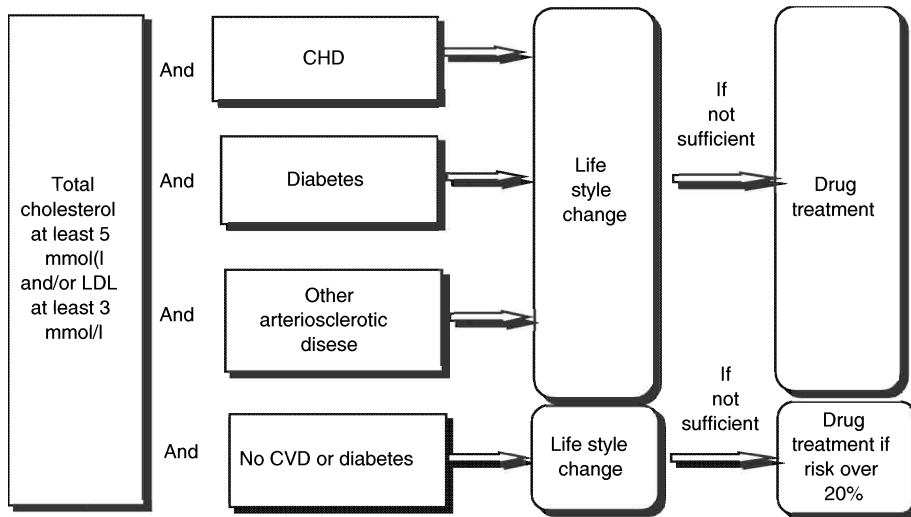


Figure 3. A simplified version of the 1999 Swedish guidelines on treatment of hyperlipidemia.

In the Swedish adaptations of the joint European guidelines the role of diabetes mellitus type 2 has been stressed and assigned the same status as CHD and other arteriosclerotic manifestations.

Just recently, new recommendations were published (Läkemedelsverket 2003). One modification is that the indication for pharmacological treatment can be extended to individuals with a total cholesterol value above 3.5 mmol/l, based on results from the HPS study (Heart Protection Study Collaborative Group 2002).

New European guidelines were released in 2003 (Third Joint Task Force of European and Other Societies on Cardiovascular Disease Prevention in Clinical Practice 2003). The focus is still on the risk estimate but the outcome is now defined as a fatal cardiovascular event rather than any kind of CHD. The reason is that the aetiology of CHD and other CVD is similar, and that intervention trials have shown effects not only on CHD but also on stroke and peripheral artery disease. The cut-off value for the 10-year risk is accordingly lowered, and has been set at 5% instead of 20%. The risk estimate system SCORE (Systematic Coronary

Risk Evaluation) is based on a number of prospective European studies and allows for different estimates for high- and low-risk countries in Europe.

While the Swedish guidelines on treatment of hyperlipidaemia are close to the joint European recommendations, it can be noted that the US guidelines are more aggressive in the sense that they promote lower target levels for LDL-cholesterol. Another difference is that they suggest a differentiation of treatment goals according to the level of risk (Smith 2001).

GPs' Use of Guidelines on Lipid Management

Awareness of the guidelines on CHD prevention and acceptance of these guidelines seem to be high according to telephone interviews with GPs in five different European countries (Hobbs 2002). A minority believed that the guidelines were largely implemented. Forty-three percent of the physicians said that they rarely or never used risk calculation charts and only 13% reported that they always did. In a sample of Danish GPs almost 60% had never heard of the European risk chart (Thomson 2001 a). Another type of evidence regarding problems in using or adhering to cholesterol guidelines comes from comparisons between data on actual prescriptions and what the guidelines suggest. A survey of medical records in the US (Abookire 2001) indicated an overuse of statines in the primary prevention situations compared with the US guidelines. On the other hand, about 70% of the patients with CHD were not on statine treatment and, among these, 88% were judged to be undertreated in comparison with the US guidelines.

Doctors' Judgements and Decisions

In their management of individual patients with elevated cholesterol values and other risk factors for CVD, GPs have to collect relevant information from the patient and the medical record. They also have to evaluate and interpret symptoms, signs and findings. They must decide whether to order further testing and what kinds of advice should be given to the patient. In the decision to recommend drug therapy or not, which is the focus of the present thesis, the decision-maker should have some idea of the effect of pharmacological treatment and the level of risk for an individual patient. If algorithms for risk estimates are not used regularly, which seems to be the case (Hobbs 2002), it is relevant to assess the ability of GPs to make intuitive risk estimates. A second and related set of questions concerns GPs' judgement strategies, which we conceptualise here as the relative importance of different risk factors in determining the inclination to initiate drug treatment.

In order to describe the judgements made by doctors, these two approaches, risk estimations and judgement strategies, have used statistical models. This

“structural” approach to the study of decisions can be contrasted to the “process” approach (Svenson 1979, Montgomery 1983, Crozier 1997), with an interest in studying the thought processes leading to a decision, with emphasis on the temporal dimension, the use of decision rules and how the mental representation of the decision alternatives changes over time. The different approaches that have been used in this thesis are described below.

Risk Estimates

The perception of a risk includes a cognitive component as well as an emotional reaction (Sjöberg 1998, Brehmer 1987). The cognitive component of risk perception involves the assessment of the probability of an event (Slovic 1987). Such assessments seem to be based on mental strategies, or heuristics, which are ‘short-cuts’ to reduce the complexity of making probability estimates (Tversky 1974, Poses 1985, Redelmeier 1995). At the same time, they can lead to systematic errors in judgement, so-called cognitive biases. Many of the heuristics seem to be relevant for probability estimates in medical decision-making (Dawson 1987). For example, the availability heuristic can mean that a diagnosis that easily comes to mind is also considered as more probable. Another example is the representativeness heuristic, where the probability estimate is influenced by how similar a case is to a diagnostic prototype.

Risk perception and risk estimates as special cases of probability estimates have been studied in a variety of fields. It seems difficult to draw any general conclusion about people’s ability to make risk judgements without taking the context of the judgements into account (Brehmer 1987, Mellers 1998). One possible generalisation is that there seems to be a tendency for small risks to be overestimated and high risks to be underestimated (Lichtenstein 1978).

A relatively small number of studies have been carried out in which the doctors’ task is to estimate future risk for cardiovascular disease based on case vignettes with various patterns of risk factors. The “true” risk has generally been calculated from the Framingham equation. Grover (1995), Friedman (1996), and Pignone (2003) demonstrated an overestimation of risk. In a study by Meland (1994), on the other hand, the GPs showed a tendency to underestimate risk, at least for high-risk patients. Thus it has not been determined whether overestimation or underestimation is more accurate as a description or under what circumstances this is the case. Schulman (1992) showed that physicians tended to overestimate the probabilities for patients with low probability for disease and to underestimate when the probability was high. Tape (1989) showed that probability estimates did not differ significantly between students and residents.

The risk reduction following drug treatment can be expressed in different ways. In intervention trials the main results are usually expressed as the relative risk, in which the outcome with active treatment is compared with the outcome with placebo. The measure might be misleading, as an impressive relative risk reduction may stem from very small changes in absolute risk if the initial risk is low. Therefore, the presentation of risk reductions is often complemented by the number of individuals who have to be treated, the numbers needed to treat or NNT, for a certain period of time to avoid one outcome. NNT is the inverse of the difference in absolute risk between the treatment group and the placebo group (Cook 1995).

It has been shown that different ways of presenting risk can affect doctors' inclination to treat elevated cholesterol values (Bucher 1994). When the results from clinical trials were presented as NNT instead of relative risk reduction, the doctors were less inclined to initiate pharmacological treatment. Cranney (1996) showed that GPs were most influenced by the relative risk reduction, which was also the case for the public (Edwards 2002).

Patients seem to underestimate their own risks compared with objectively estimated risks (Avis 1989, Niknian 1989).

Weighting of Evidence, Clinical Judgement Analysis

Medical judgements and decisions often have to be based on information that has an uncertain relation to the underlying disease. For a certain symptom or combinations of symptoms there can be different underlying diseases, and the same disease can give different patterns of symptoms and signs from patient to patient. In other words, there is a probabilistic relation between the underlying condition and the evidence. The clinical findings often have to be weighed together to produce an overall judgement or a decision. For example, are the symptoms and signs taken together enough to justify antibiotics for a suspected pneumonia? The theoretical basis for the study of people's adaptation to an uncertain environment is Brunswick's probabilistic functionalism (Cooksey 1996). When applied to medicine, the approach has often been termed Judgement Analysis or Clinical Judgement Analysis (CJA) (Kirwan 1990). In a typical experiment, a doctor is asked to judge a number of patient cases within a specified medical domain. The information in the case vignettes varies systematically with regard to a number of variables, e.g. degree of symptoms, values on laboratory tests, patient's age and so on. These variables are called cues.

The judgement or decision in a given case is typically a value on a graded response scale (e.g. the probability that the patient has the diagnosis in question).

Not only will a cue have be correlated with the underlying state of the patient, but it will also be correlated with the judgement made by the physician. This symmetrical model has been termed the lens model and is schematised in Figure 4.

To exemplify from the medical field, the lines on the left side of the lens depict the relations between a disease and the different symptoms, laboratory values and so on. On the right side of the lens are the different weights that the doctor assigns to different cues.

Lens model studies are closely linked to the statistical model used, which is typically linear multiple regression. Each patient variable is assigned a regression weight, which is supposed to represent the relative importance of this variable in determining the judgement. In some CJA studies the true state of the patient and its correlation to the cues are not known, and the study concerns the right side of the lens. The pattern of statistical weights for a given doctor is often called judgement policy or decision policy. Much of the research on the medical application of the lens model has concerned the study of such policies, e.g. the extent to which there are differences between different doctors. Some of the findings (Wigton 1996) are that there is surprising variation between different physicians, and that this variability seems to be present both in experts and in novices. The divergence in policies has been demonstrated in various disease groups such as, for example, rheumatology (Kirwan 1983 a and b), psychiatry (Fisch 1981), heart failure (Skånér 1998, 2000) and lipid management (Evans 1995). Another finding in these studies is that cues that are given a substantial weight by the doctors differ from those recognised in textbooks. A third recurrent finding according to Wigton (1996) is that the number of cues used is surprisingly few, often only two or three, even if many more cues are present.

The correspondence between cues that are actually used and cues the doctors think they have used has often been found to be low (Brehmer 1988). It has been hoped that giving feedback to doctors about the outcomes of their decisions would improve judgements. The results are quite complex, but it seems that simply providing the doctor with the correctness of each decision does not improve performance on later sets of cases. If, however, the feedback is provided as so-called cognitive feedback showing the decision weights used by the doctor and contrasting them with the best possible cue weights, this may lead to better decisions (Klayman 1988).

Evans (1995) used the CJA paradigm to study cholesterol treatment. Thirty-five GPs rated 130 imaginary cases presented by a computer. Thirteen cues were examined and the cases were compiled by randomly generating combinations of

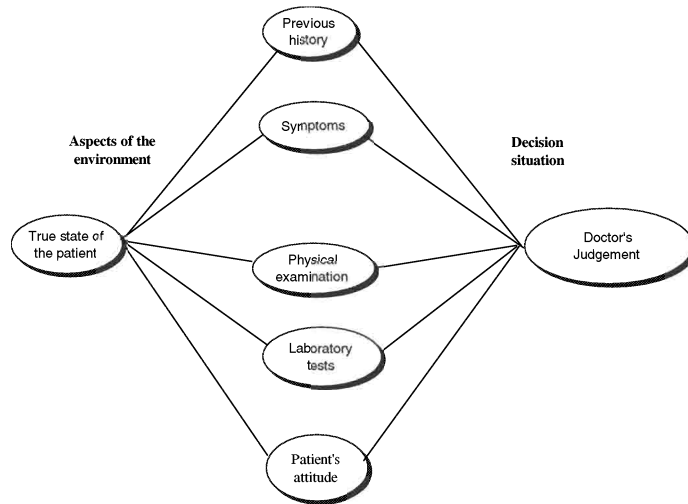


Figure 4. A simplified representation of the lens model applied to a medical judgement task.

cue values. It was shown that doctors overestimated the number of cues that had actually influenced their decisions. It was further shown that among the traditional risk factors for CHD, only diabetes had a significant effect on the prescription decision.

The CJA approach has been questioned from a number of points of view. Its clinical validity may be questioned on the grounds that in real clinical settings information is seldom available all at once. Instead, clinical work is characterised by a gradual gathering of information with resulting modification of the judgement. Also, alternative models have been presented to describe the results of CJA studies, assuming that decisions are not necessarily characterised by a linear compensatory integration of several sources of information but instead are based on simpler rules and in some situations only on one cue (Dhami 2001). There has also been controversy about the extent to which the lens model should claim to have anything to say about the judgement processes beyond describing the mathematical relations between the presented information and the decisions (Doherty 1997).

Decision-Making as Cognitive Processes

The statistical models for the study of decisions describe the relation between information and judgements or decisions without directing much attention to the psychological processes on the way to the decisions. They are sometimes called “black-box” investigations as they concern only the structure between input and output (Elstein 1978). Several issues related to medical decisions are not so easily addressed using these approaches. They include the following: Which hypotheses about diagnosis are generated and tested? To what extent are preliminary decisions about therapy or diagnosis generated and how are they evaluated when new information is provided? What decision rules and medical knowledge seem to be applied in the decisions? How should the medical and procedural knowledge, or competence, be described for different levels of training and medical expertise?

An alternative approach to statistical modelling of the relation between information and decision is a process approach where the focus is on the decision process as such, and so-called process-tracing techniques are used (Svenson 1979). Process explanations of decision-making emphasise the temporal dimension of decision-making, decision rules for selection and comparison of decision alternatives, and the idea that decision-makers change their cognitive representation of the decision alternatives in order to reach a decision (Crozier 1997).

Different methods have been used to follow how the decision develops over time, such as tracking of eye movements, recording of the information demanded by the decision-maker (Svenson 1979) and repeated ratings (Dahlstrand 1984). The method used to describe the sequence of thoughts behind decision-making is typically think-aloud (Ericsson 1980, 1999). The subjects are instructed to say aloud all their thoughts while performing a task. The underlying theoretical formulation is that the process of thinking is “..a sequence of states, each state containing the end products of cognitive processes, such as information retrieved from long-term memory, information perceived and recognised, and information generated by inference” (Ericsson 1999, p xiii). Individuals are supposed to be able to report accurately what is in their focus of attention. However, they are less accurate in reporting the underlying processes, e.g. how they arrived at a certain solution to a problem. For the think-aloud method to be a valid reflection of the thought processes involved in a task, the subjects should therefore not be asked to explain their decisions or inferences, but just to say aloud what is in their focus of attention at the moment. Explaining would introduce additional thinking that is not relevant to the task, and thus decrease the validity of the method.

The verbal reports are usually audio-taped and thereafter transcribed into a written form and segmented into statements, roughly corresponding to sentences. The set of statements is thereafter coded using a protocol that is constructed beforehand and that is based on a theoretical analysis of the task.

The validity of the think-aloud method has been questioned (e.g. Nisbett 1977). One source of invalidity is the risk that verbalisation may change the cognitive processes we want to study in terms of their content or processing time. Another source of invalidity is that the protocol may not reflect the underlying cognitive processes that are involved in solving the task. According to Ericsson and Simon (1980, 1999), thinking aloud while performing a task often lengthens the time to complete the task, but does not seem to change the accuracy of task fulfilment or the cognitive processes. However, in studies that involve instructions that may cause the subjects to try to explain their thoughts, this general rule does not hold. If the task contains non-verbal information that has to be coded verbally, this could decrease the validity of verbal reports (Schooler 1993).

Think-Aloud Technique Applied to Medical Decision-Making

In a series of well-known early studies on medical decision-making using think-aloud technique, Elstein and co-workers (1978) examined the generation of diagnostic hypotheses. Their conceptual framework was the medical diagnosis as a hypothetico-deductive process. One main conclusion was that a small number of tentative diagnostic hypotheses are generated very early in the patient encounter. In about 90% of consultations that concerned setting a diagnosis, the first diagnostic hypothesis was activated as soon as the patient had phrased his or her main complaint. Another conclusion was that the most common error is "overinterpretation", meaning that new evidence is more often falsely regarded as supporting the tentative hypothesis than correctly interpreted as speaking against it. A third conclusion was that there was no evidence of a general difference between experts and novices in terms of a general problem-solving skill. Rather, differences in competence between different levels of expertise seem to be best described by differences in the quality of the knowledge representations. A more efficient knowledge base in terms of content and structure could then lead to better diagnostic hypotheses on the part of the experts.

The concepts of knowledge organisation and knowledge structure as the important dimensions for describing medical competence have been termed 'script' or 'illness script' (Schmidt 1990, Custers 1996 and 1998, Charlin 2000). An approximate definition of an illness script is the mental organisation of a disease or a medical procedure or some other limited medical field (Schmidt

1993). According to this view, the generation of a diagnosis can be described as the activation of a script. The biomedical details of the disease with its causes are not very important in most routine diagnostic situations, and the scripts are instead characterised by so-called enabling conditions (background factors like age, sex, and previous diseases). Less experienced physicians have less training in recognising patterns of enabling conditions and corresponding diagnoses and they therefore have to rely on biomedical reasoning.

The difference between different levels of expertise has also been the concern in a large number of studies with think-aloud technique by Patel and co-workers. By contrasting the verbal protocols from experts and novices while solving difficult diagnostic tasks they have suggested different diagnostic strategies depending on their level of clinical experience (Patel 1986, Arocha 1993). Hypothesis generation is supposed to be driven by forward reasoning among experts, meaning that data serve to trigger a hypothesis. Backward reasoning is, on the other hand, characterised by generating a hypothesis first and then determining whether data conform to the hypothesis. More experienced physicians seem to generate a smaller set of initial hypotheses and to be more efficient in testing them against new evidence (Joseph 1990, Allen 1998, Kushniruk 1998).

Most of these studies have focused on the diagnostic processes in solving quite difficult cases. Few think-aloud studies have addressed the decision processes behind treatment decisions. In a study of the treatment of pulmonary lung embolism, Kushniruk (1995) coded the think-aloud protocols for hypothesis generation and for activities in the decision situation including framing of the situation. The results suggested that experts tended to focus on developing more refined situation analyses of the decision problem and were less often engaged in ordering tests as compared to the less experienced doctors.

Process Theories on Decision Making

In the think-aloud studies that will be reported here, two process theories on decision-making have been used to make predictions about the decision process, the Search for Dominance Structure Theory and the Differentiation and Consolidation Theory. They are two examples of theories that assume a restructuring of decision alternatives in order to favour one alternative before making the choice, so-called biased predecision processing (Brownstein 2003). The theories both suggest that a preliminary alternative for the choice be selected early in the process. In our case this would mean that the doctor makes a preliminary decision on whether to recommend a drug or not before all the information is presented. In his Search for Dominance Structure Theory,

Montgomery (1983, 1989) suggests that the decision-maker tries to structure the decision situation so that the rule of dominance can be applied: one decision alternative has no disadvantages and at least one advantage compared with the others. The theory predicts that the evaluative difference between the to-be-chosen alternative and competitors will increase as the decision process proceeds. The same prediction can be made from Svenson's (1992, 1996) Differentiation and Consolidation Theory, although from a partially different rationale than search for dominance. In Study III the prediction was made from the theories that the GPs who finally chose drug treatment for a certain case would show an increasing inclination over time for this choice, whereas the reverse would be true for GPs who decided not to prescribe a drug.

Aims

The general aims of the five studies were to increase our understanding of how Swedish GPs make decisions concerning whether to initiate drug treatment or not for individuals with elevated cholesterol values. Increased knowledge in this area is supposed to enhance our possibilities to improve guidelines and teaching within the area of managing risk factors for cardiovascular disease and to design training and education. A related purpose was to examine the usefulness of different methodological approaches to judgements and decisions within this medical field.

More specific aims were:

- to describe GPs' strategies in terms of which factors influence their decisions to prescribe lipid-lowering drugs, and how their judgements agree with guidelines on the management of hyperlipidaemia (Study I)
- to evaluate GPs' ability to make intuitive risk estimates for individuals with high cholesterol values, and to compare their estimates with those of less experienced clinicians, i.e. medical students in their final year of training (Study II)
- to examine the validity of think-aloud reports, compared with traditional rating data, for describing the decision process over time and to examine whether the addition of a rating task would affect the think-aloud data (Study III)
- to describe, by using think-aloud data, how GPs evaluate different kinds information about the cases in their decisions on drug treatment (Study IV)
- to analyse how GPs' knowledge of guidelines corresponds to their decisions on patient cases and what the arguments are for their decisions (Study V).

Methods

Patient Cases/Case Vignettes

Written patient cases were used in all studies. For Study I, 40 patient cases were selected consecutively from the laboratory records of a primary health care centre. Patients were selected who had a total cholesterol value of at least 5.5 mmol/l (which was the lowest value at which drug treatment should be recommended according to the 1995 Swedish guidelines that were in use in 1997 when the study was conducted), and who were not already on lipid-lowering medication. The 40 most recent patients, based on their occurrence in the records, who fulfilled the criteria were included. The 40 case vignettes were presented in a paper format and sent out as a postal questionnaire.

In Study II, six patient cases were chosen from the population of 40 cases in Study I. They were chosen to represent a spread of risk according to the Framingham equation. Four cases represented primary prevention and two represented secondary prevention. While these six cases were authentic, a second set of four cases was non-authentic and was constructed to examine the effect on risk of the degree of cholesterol elevation and the effect on risk of lowering the cholesterol level by drug treatment. Altogether, then, there were 10 case vignettes. The central characteristics of the cases along with the treatment recommendations according to the 1999 Swedish guidelines on treatment of hyperlipidaemia are presented in Table 2.

Studies III-V were based on the same six patient cases. As in Study II, the set that was used was chosen from the original set of 40 cases from Study I. The cases were chosen to represent one secondary prevention case, one case of diabetes, and four primary prevention cases with a spread in the 10-year risk. The details of the cases are described under Results and in Table 3.

Participants

In Study I, 60 GPs from the southern Stockholm area were randomly selected from an available list of 90 specialists in family medicine.

In Study II, the GP group was a random sample of 200 from a list of all practising GPs in Sweden. The student group was all 73 students who joined a seminar in family medicine in their final year of medical school.

Studies III-V were based on the same original data. Twenty doctors from the southern Stockholm area participated. They were selected so as to be relatively evenly distributed across different districts and between sexes, but the selection was not randomised. A total of 36 doctors were contacted by telephone. Twenty-four agreed to participate, but of these, four later cancelled their participation before the session.

Design, Procedure and Data Analyses

Study I

This study used the CJA approach to compare the relative weights given to different kinds of information (“cues”) about the patient cases in the judgements concerning whether to initiate pharmacological treatment or not.

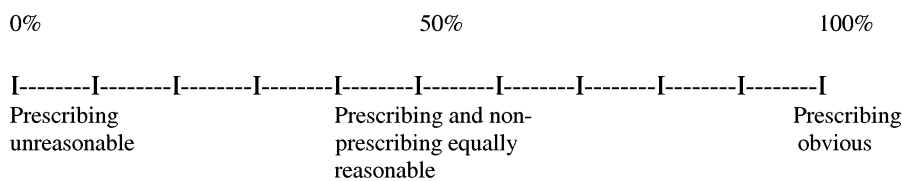
The 40 patient cases were sent out as a postal questionnaire to the GPs’ home addresses, with two reminders if necessary. The task of the participating doctors was to rate their inclination to drug treatment for the different cases. The written instructions specified that all patients had at least a mild elevation of cholesterol and that an initial attempt at non-pharmacological treatment had been made for at least six months.

The following is an example of a case vignette with the response scale.

“A 64-year old woman with hypertension. Does not have diabetes. Suffers from coronary heart disease but has no other atherosclerotic disease. Non-smoker.

Laboratory values: cholesterol 6.0 mmol/l. LDL 4.1 mmol/l. Triglycerides 1.9 mmol/l.

Try to grade, in percentages, how willing you would be to prescribe a cholesterol-lowering drug for this case.



Put a mark on the line to show your standpoint.”

At the end of the questionnaire, the doctors were presented with a list of the nine cues used and were asked to rank the three cues thought to be of most importance for their decisions. They were also asked whether they had used any written guidelines or other written material as assistance.

In the multiple linear regression model, the judgements were used as the dependent variable and the different cues as independent variables. The judgement was made on a visual analogue scale (VAS) of 0-100%. The end-points were defined as 0%, prescription totally unreasonable, 100%, prescription obvious, and 50%, prescription and non-prescription equally reasonable.

A so-called representative design was used. The sampling of authentic patient cases aimed at giving a distribution of patient characteristics that were representative for the group of patients with elevated cholesterol values that GPs meet. An alternative would have been to select a random sample of all possible combinations of cue values. While this technique holds down the correlation between cues (a so-called orthogonal design), which in turn gives the method a greater power, the disadvantages are cue combinations that are rarely seen in real life, and thereby there is a risk of a lower face validity of the investigation.

The cues used (the independent variables in the regression model) had to be limited in order not to generate too many case vignettes. The cues were selected to be compatible with the current guidelines on lipid treatment. Nine cues were used: age, sex, hypertension, diabetes, CHD, smoking, total cholesterol, LDL and triglycerides. As total cholesterol and LDL were very highly correlated, only the former variable was used in the model. Age and triglycerides were dichotomised in the regression model.

Each doctor's judgements were subjected to multiple regression. The relative importance of each cue was defined as its regression weight. Statistical significance was $p < .05$. For each cue, the mean regression weight over the 38 doctors was calculated. For each doctor, the square of the multiple correlation derived from the regression equation (r^2) was calculated and this measure is equal to the proportion of judgement variance that is explained, i.e. how well the regression model describes the individual doctor's judgements. Student's t-test was used for differences between groups in mean ratings and mean age, Pearson's r was used to estimate associations between continuous variables, and the chi-square test was used for associations between dichotomous variables.

Study II

The principal aim of this study was to describe how well GPs and medical students can make subjective estimates of the 10-year risk of coronary events for patients with hypercholesterolaemia.

The task of the GPs and the students was to estimate the risk of future CHD within 10 years for each case and to decide whether to recommend a lipid-

lowering drug or not. The risk was defined as the number of persons, out of a hypothetical group of one hundred persons with the same risk factors who would have CHD within 10 years. There is evidence that information presented as frequencies rather than as probabilities, which are more common, facilitates risk judgement (Gigerenzer, 1996).

The following is an example of a case description.

“Case 1. 41-year-old women with normal blood pressure (80 mm Hg) and without diabetes. No coronary heart disease or other atherosclerotic disease. Smokers.
Laboratory values: Total cholesterol 7.2 mmol/l. LDL 5.3 mmol/l. HDL 1.2 mmol/l.
My answer _____ persons would have CHD within ten years

Would you recommend a lipid-lowering drug? No Yes”

The questionnaire was sent out to the GPs’ work addresses, while the students were asked to fill in the questionnaire in connection with a seminar.

The 10-year risk was calculated using an equation described by Anderson (1991), which is based on the original Framingham cohort combined with the second-generation study population. A parametric regression model was used with different equations for systolic and diastolic blood pressure. A comparison was made with a prediction equation based on a more recent Swedish material, the Malmö Preventive Project (Berglund 1996). The variables selected to describe the patient cases were compatible with both equations. LDL values were specified in the case descriptions, as LDL plays a central role in the 1995 Swedish guidelines, but LDL was not included in the risk prediction equations. Left ventricular hypertrophy is specified in the Framingham-based formula but was excluded (the value was set to zero in the equation), as reliable information was difficult to retrieve from the records.

The Framingham-based equation was used for the calculation of 10-year risk for the primary prevention cases. For the Swedish equation, the risk estimations were made by the “Risk score 99” computer program, available as a floppy disc from the Hässle Medical Company. The patient characteristics and resulting risks are shown in Table 2 along with the suggested treatment according to the 1995 Swedish guidelines and, as a comparison, the 1998 joint European guidelines.

Student’s t-test was used for differences in age between responders and non-responders among the GPs, the chi-square test was used for the association between sex and response rate (GPs), the Mann-Whitney test was used for the difference between medians (GPs and students), and Freedman’s test was used for comparing the two groups in terms of response patterns across cases. Spearman’s

rank correlation was used to describe the correspondence of the individual risk estimations with the estimations from the prediction equations.

Studies III-V

These studies were based on the same experiment.

One of the purposes of the first of the three studies (Study III) was to investigate whether think-aloud protocols would be affected if a simultaneous task of using a rating scale were added. The 20 doctors were therefore randomised into one of two groups, a think-aloud condition and a think-aloud+rating condition with 10 doctors in each.

The 20 doctors received the same six patient cases and in the same order so as to reduce the variance due to differences in case order and thereby facilitate the comparison between the two randomised groups. The different kinds of information were segmented temporally in the same way for all the six cases. Table 1 shows a complete example of one case.

Table 1. Example of a case presentation (Studies III-V).

Screen	Information
1	The patient is a 67-year-old woman whose recent cholesterol value was 7.3 mmol/l. She has had the diagnosis hypercholesterolaemia for two years. She has been given advice concerning diet but she has not been prescribed a cholesterol-lowering drug. Her cholesterol value has decreased from an initial value of 7.8 mmol/L
2	The patient has been on medication for hypertension for 10 years (Seloken ZOC* 50 mg and Plendil** 5 mg). She is now on a visit to check her blood pressure and hypercholesterolaemia.
3	The patient has no other known diseases apart from osteoarthritis of her knees. Her mother suffered from hypertension and reached the age of 84 years.
4	The patient is a non-smoker. She very seldom drinks alcohol. She does not exercise on a regular basis but she is fond of taking walks.
5	Physical examination: Good general condition. A few kilograms overweight. Blood pressure 145/75. Heart and lung auscultation normal.
6	Laboratory values: Total cholesterol 7.3 mmol/L. LDL 5.4 mmol/L. HDL 1.2 mmol/L. Triglycerides 1.6 mmol/L. TSH, creatinine and liver-function tests were normal.
7	Would you prescribe a cholesterol-lowering drug for this patient? Yes No

The GPs controlled the shift to a new screen by a mouse-click. They were instructed that authentic cases of hypercholesterolaemia would be presented and that their task was to say aloud all their thoughts about the cases and that each case would end with the question as to whether or not they would prescribe a drug for this patient. If the doctor was silent for more than 10-15 seconds he or she was reminded to talk aloud. After the six cases, they were asked to describe in their own words how they usually reason in their practice when drug treatment for elevated cholesterol values may be indicated. The resulting protocols have been labelled Free-Report Protocols to be distinguished from Think-Aloud Protocols, which refer to the processing of the different patient cases. The following data were collected.

- Time spent on each screen, which was automatically logged.
- Decision whether to prescribe or not for each case.
- Ratings (10 of the 20 doctors). The doctors moved a pointer along an analogue scale and the ratings were recorded in 11 steps, 0,10,20...100%.
- Think-Aloud Protocols. The tape-recorded sessions were transferred into a written word-by-word format. The protocols were segmented into statements (the smallest meaningful unit according to the coding scheme used). The statements were coded for three aspects. First they were coded into one of 10 cognitive categories, exemplified by attention, evaluation and explanation. Second, each statement was coded with regard to its directionality in relation to the decision task (to prescribe a drug or not). Third, each statement was coded with respect to the information referred to, e.g. the patient's age or LDL level (for use in Study IV). The Think-Aloud Protocols were also coded (for use in study V) for cause-effect relations with relevance to two aspects of the guidelines on hyperlipidaemia–secondary prevention and diabetes. They were finally coded for arguments to prescribe or not for the individual cases.
- Free-Report Protocols (Study V). As was done in the Think-Aloud Protocols, the Free Report Protocols were segmented into statements. These were then coded for cause-effect relations (Axelrod 1976). In the present context, causes were mainly variables describing patients, e.g. their age, laboratory values, smoking habits and so on. A cause variable could be a single variable or a combination of different cause variables. The original coding resulted in 34 different single cause variables and 39 different combinations of two or three. The number of effect variables was 23, but only those three were used that were judged to be directly relevant to the task in the instructions, i.e. to describe their thoughts about drug prescription in lipid management. The three effect variables were: Usability (indicating the degree to which a certain kind of information was useful for the decision situation), Probability (indicating that a certain cause variable influenced the probability for

the doctor to prescribe a drug), and Implication (indicating that the indication for prescription was definite or almost definite).

In Study III the mean values for ratings and the mean values for directionality were calculated for each screen for the group of 10 who delivered both kinds of responses. ANOVA was used with rating or directionality, each one separately, as the dependent variable. Decision (Yes/No) was used as a between group variable and screen (1-6) as a within-group variable. In all ANOVAs with repeated measures the degrees of freedom were adjusted according to Greenhouse-Geisser Epsilon in SPSS.

How well the decisions (Yes/No) could be predicted by the two response measures, ratings and directionality, was also assessed by signal detection theory (Tanner 1954). This theory makes use of four outcomes of a decision. To simplify, and to apply the theory to a medical setting, a hit is defined as the case when the doctor judges that a certain condition is present or will occur and this turns out to be true, a false alarm is when it turns out not to be present or not to occur. The corresponding outcomes when the doctor's judgement is that the condition is not present are a miss and a true rejection, respectively. In Study III the doctors' judgements were the successive responses to the presented information, i.e. the ratings and the directionality of the statements. The outcomes were if the final decision about the case for a certain doctor was "Yes" or "No". Ratings above the midpoint (50%) or statements with a positive directionality were defined as "hits" if the decision was "Yes", and as false alarms if the decision was "No". The difference between the z-transformed proportions of hits and the z-transformed proportions of false alarms defines a sensitivity index, d' , which is a measure of how well the ratings and think-aloud data could predict the final decision at different phases of the decision process.

In Study IV the importance of different kinds of information about the patient cases was evaluated. Eleven information categories were used: Cholesterol, LDL, HDL, weight, smoking, CHD, diabetes, hypertension, heredity, sex and age. Importance was defined as the number of evaluative statements (i.e. with a positive or negative directionality).

In an ANOVA with repeated measures the number of evaluative statements was the dependent variable and the three within-group variables were Direction (Positive/ Negative), Information Category (1-11) and Case (1-6). The six cases were analysed one at a time in ANOVAs with Decision (Yes/ No) as a between-group variable and Case and Information Category as two within-group variables.

In order to identify subgroups of doctors with similar patterns as to how they evaluated different information categories, a hierarchical cluster analysis was performed according to Ward's method (SPSS). Twenty-two variables (each of the 11 information categories and each with a positive or negative directionality) and 20 participants were entered into the analyses.

Results

Study I

Response Rate

Thirty-eight doctors (63%) answered the questionnaire. Four doctors stated that they had used some form of decision support when they answered the questionnaire.

Mean Ratings

The total mean rating on the scale from 0% to 100% was 54%. The individual doctors' mean rating varied from 28% to 78%. There was no significant correlation between mean rating and the doctor's age or years of experience.

Regression Weights

The mean regression weights for the group of 38 GPs are shown in Figure 5.

CHD had the highest mean weight at 26, i.e. the presence of CHD raised the VAS ratings on average 26% with all other cues assumed to be constant. Next in importance was cholesterol. All the cues except age had a positive weight (a positive weight for sex indicates a greater tendency to prescribe drugs for male patients, other variables being equal). The difference between doctors in terms of their patterns of regression weights (strategies) was substantial. For 28 of the 38 doctors, two or three of cues were statistically significant at the .05 level. Eight doctors appeared to have used four or more cues, while two doctors used one or none. The mean proportion of explained variance (r^2) was .62 for the whole group of GPs (range .18–.90).

Relation Between Actual and Stated Strategies

For each doctor, an "objective" ranking of the cues in terms of the sizes of the regression weights was established. The resulting ranks could be compared with how the doctors thought they had used the cues ("subjective" ranking). CHD and cholesterol were highest in both objective and subjective importance. Hypertension and diabetes were significant in the regression equation for less than



Figure 5. Mean regression weights for the 38 GPs.

a third of the doctors. It can also be noted that nine doctors (24%) did not use CHD as a significant cue. One of these doctors used no cue at all according to our statistical definition. Of the remaining eight, one had ranked CHD as number one and one as number two in subjective weight. As a comparison, among the 27 doctors who gave CHD a significant weight, 16 had ranked it as number one and another six as number two in subjective importance. In other words, the GPs were generally aware of their strategy of using or ignoring CHD. The doctors who used CHD were younger (47 years) than the doctors who did not (52 years), $p < .05$. The groups did not differ significantly with respect to sex or r^2 .

Comparison with Guidelines

Twelve of the 40 patients had CHD and should be recommended pharmacological treatment according to the 1995 Swedish guidelines. The mean rating for this group was 65%. Two of the cases had mean ratings below the midpoint (30% and 44%). Both had lower cholesterol levels, 5.7 mmol/l and 5.9 mmol/l, than the average for the whole subgroup, 6.4 mmol/l. In addition, one of these two patients was the oldest in the set of 40 (78 years). For 11 patients, pharmacological treatment “should be considered” according to the same guidelines. The mean rating for this group was 62%. For the remaining 17 patients, pharmacological treatment was not in accordance with the guidelines, and the mean rating was 40%.

Study II

Response Rate

Of the 200 GPs, 42% returned the questionnaire and 35% gave complete answers on both treatment decisions and risk estimates. There were no significant differences between responders and non-responders regarding age or sex. Only nine GPs (11%) stated that they had access to decision support for risk calculation.

Of the 73 students, 57 (78%) answered the questionnaire.

Risk Estimates for Different Cases

Figure 6 shows boxplots for the distribution of ratings for each of the cases. The Framingham risks are marked for the primary prevention cases (upper figure). The risks were underestimated by both GPs and students, except for case 1, representing the lowest risk. The lower figure shows the corresponding results for the two CHD cases, for which the Framingham risk could not be calculated. However, it has been suggested (Wood 1998) that the presence of CHD would increase the risk by at least one category in the risk chart. The resulting 10-year risk would be approximately 35% for case 2 and 45% for case 4. This indicates a tendency to underestimate risks for secondary prevention cases as well.

Isolated Hypercholesterolaemia

The reference case with a cholesterol value of 5.0 mmol/l (no. 7) had a 10-year risk of 12%. The absolute risk increase with an assumed increase in cholesterol to 6.5 mmol/l (case no. 8) was 4%. The average difference in risk estimates between the two cases was 3% for GPs and 5% for students. When the cholesterol value was assumed to increase from 5.0 to 8.0 mmol/l, the resulting increase in Framingham risk was 7% and the corresponding mean risk increase for GPs was 10% and for students it was 11.5%. Lowering the cholesterol level from 8.0 mmol/l to 6.5 mmol/l with pharmacological treatment was estimated by both GPs and students to result in the same risk as initially having a level of 6.5 mmol/l.

Accuracy

Accuracy was defined as the correlation between the subjective and the algorithm-based risk estimates. The Framingham equation and the Swedish equation gave exactly the same ranks to the primary prevention cases. For these cases the median rank correlation was 0.89 for GPs and 0.80 for the students, and the difference was not significant. If we include secondary prevention cases (using the Swedish

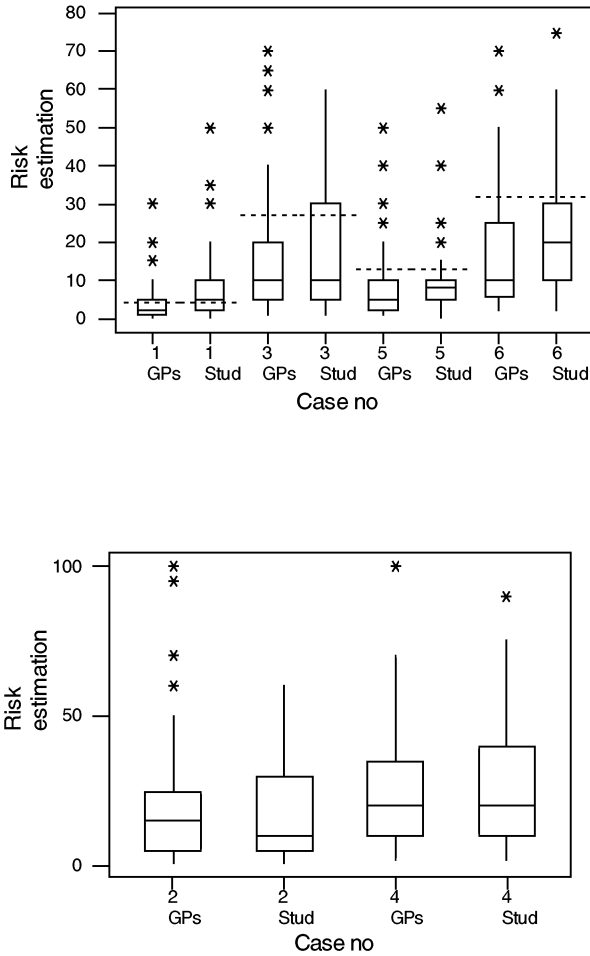


Figure 6. Ratings by GPs and students as boxplots of primary prevention cases (upper figure) and secondary prevention cases (lower figure). Dashed line denotes the Framingham risk. The lower and upper limits of the boxes represent the third and first quartiles of the observations and the lines in the boxes are the medians. “Whiskers” are drawn to the highest and lowest values. When these are extreme, they are marked with asterisks.

equation), the median rank correlation for the GPs was significantly higher than for the students (0.79 vs. 0.68, $p < .05$).

Treatment Decisions

Table 2 shows that the case with the highest proportion of respondents willing to treat was no. 9, representing isolated hypercholesterolaemia (8.0 mmol/l); almost every student and three GPs out of four were inclined to prescribe a drug. On the other hand, only about half the GPs and students suggested drug treatment for the two CHD cases. More than twice as many students as GPs were inclined to treat the case with the lowest risk (no.1).

Study III

Interjudge Reliability of Coding

Two of the authors coded the protocols from the first six participants independently of each other. Reliability was calculated and defined as the percentage of statements that were coded identically. The reliability was 91% for cognitive categories and 92% for directionality.

Comparing Ratings and Think-Aloud Protocols (Validity)

The prediction was that, for a certain case, the difference between the groups who finally gave “Yes” and “No” –decisions, respectively, on whether they wanted to prescribe a drug would increase over screens. The corresponding methodological question was whether this increased differentiation would be equally or differently reflected by the two response measures, ratings and think-aloud protocols. To address these questions the directionality of ratings and statements were compared for the think-aloud+rating group. At each screen, the decision directionality for statements was defined as the total of the “+”, “0”, and “-” statements given on that screen. Figure 7 shows the most illustrative example of this gap between the groups across screens.

Statistically, the increasing gap, or the increased differentiation, corresponds to the interaction between Decision and Screen. Both kinds of data showed a tendency toward an increased differentiation between the “Yes” and “No”-groups, while the interaction was significant in more of the cases with statements than with ratings. In other words, think-aloud data were at least as valid as ratings in describing the decision process.

Table 2. Summary of the ten cases (Study II). Risk estimates are given in percentages. For each case, the GPs' and students' willingness to prescribe a lipid-lowering drug are shown in percentages and whether drug treatment is recommended by Swedish and joint European guidelines.

Case	1	2	3	4	5	6	7	8	9	10
Age	41	56	66	70	51	61	55	55	55	55
Sex (M male,F female)	F	M	F	M	F	M	M	M	M	M
Hypertension	No	No	No	No	No	No	No	No	No	No
Diabetes	No	No	Yes	No	No	No	No	No	No	No
Smoker	No	No	No	No	Yes	Yes	No	No	No	No
Total Cholesterol	7,2	5,9	6,9	6,0	6,5	8,2	5,0	6,5	8,0	6,5
LDL	5,3	4,3	4,6	4,1	4,3	6,3				
HDL	1,2	1,1	1,1	1,0	1,0	0,9				
Primary/secondary Prevention	Prim	Sec	Prim	Sec	Prim	Prim	Prim			
Framingham risk	4		27		14	33	12	16	19	16
Median risk GPs	2	15	10	20	5	10	5	7	15	8
Estimated median risk students	5	10	10	20	8	20	5	14	20	10
Percentage of GPs inclined to treat	16	39	57	49	11	68		0	73	
Per cent of students inclined to treat	41	56	54	45	19	86		33	93	
Drug treatment (Swedish guidelines)	No	Yes	No	Yes	No	Consi der	No	No	Consi der	
Drug treatment (joint European guidelines)	No	Yes	Yes	Yes	No	Yes	No	No	No?	

GPs' decision-making on drug treatment of hypercholesterolaemia

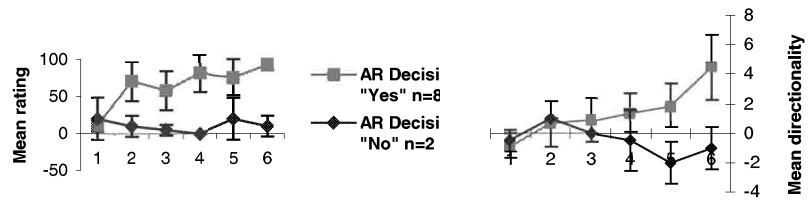


Figure 7. Directionality for ratings (left curve) and statements (right curve) for one case.

The two measures were also compared using signal detection theory. An alternative way of scoring for directionality was introduced in this analysis: the cumulative directionality at a certain screen was the sum of all “+” and “-” up to and including the screen. Figure 8 shows that the sensitivity index, d' , was almost identical for ratings and non-cumulated directionality and had slighter higher values for cumulated directionality.

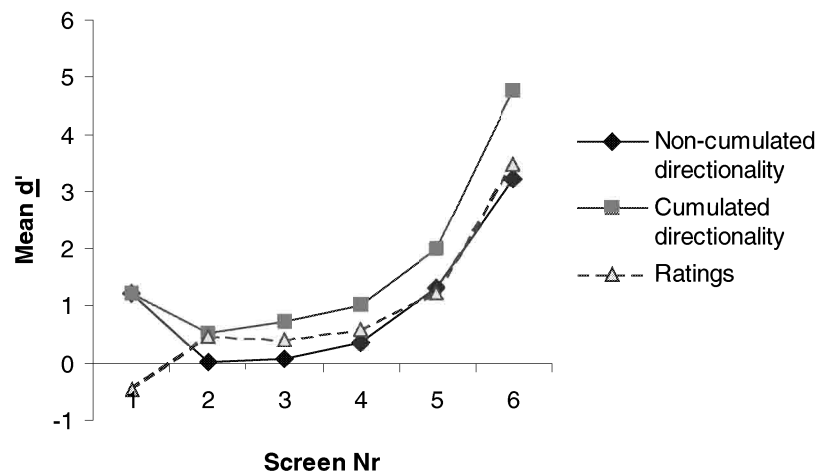


Figure 8. Signal detection theory applied to the different response measures. Mean d' for the six cases are shown for three different response measures and for the six screens.

Effects of Adding a Rating Scale

The second methodological question was whether the addition of a rating task would affect the decision process or the contents of the verbal protocols. To answer this question the two randomised groups were compared in four respects: the final decision “Yes” or “No”, the decision directionality over time, the time spent on the cases, and the contents of the think-aloud protocols.

- The final decision. The patterns of Yes/No decisions across cases were almost identical when the two groups were compared.
- Decision directionality over time. In none of the patient cases was there a significant interaction effect in ANOVA that included Instruction (think-aloud vs think-aloud+rating). These results indicate that the additional task of rating did not affect the think-aloud reports, regarding their ability to differentiate the response pattern over screens for the two decision outcomes.
- Time spent on the cases. The think-aloud+rating group spent a slightly longer time, summarised over the six cases (mean 17 min, 6s), than the think-aloud group (mean 16 min, 5 s), but the difference was not significant.
- Contents of the think-aloud protocols. The think-aloud+rating participants made significantly more statements concerning pharmacological treatment and fewer statements concerning other actions such as life-style change.

Study IV

While Study III had a methodological focus, the aim of this study was to examine how GPs use clinical information and knowledge of guidelines in their treatment decisions.

Reliability of Coding

In 94% of the statements (protocols from the first six participants) the two investigators made identical codings of the information category or categories. The inter-judge reliability was 92% for the directionality of the statements (as described in Study III).

Information Categories

The original set of 21 information categories that was used to describe the patient cases was reduced to 11. When the information categories were ranked with regard to the frequency of positive and negative evaluations, there was a great leap between overweight (frequency 14 and rank 11) and triglycerides (frequency four and rank 12). Therefore triglycerides and information categories with fewer

evaluations were excluded. The information categories that were further analysed were Cholesterol, LDL, HDL, overweight, smoking, CHD, diabetes, hypertension, heredity, sex and age.

Treatment Decisions

Table 3 summarises the information for each case as regards these 11 information categories and shows the number of doctors who decided to prescribe a drug. It also shows the recommended decisions according to the Swedish guidelines.

The total number of "Yes"-decisions and "No"-decisions was the same when summarised over patients and cases. The majority of doctors chose to prescribe for two, three or four of the six cases. No participant decided to prescribe for all six cases and one participant chose not to prescribe for any of the cases.

Importance of Information

Figure 9 shows that information about cholesterol was evaluated most frequently, both in the positive and the negative direction. Positive evaluations were more frequent than negative ones, except for sex, age, and weight. ANOVA with the number of statements as dependent variables showed significant effects of Information Category and the interaction Information Category X Case. This indicates that the different information categories were evaluated unequally often, and that the patterns of relative importance differed in the six individual patient cases.

Patterns of Importance for "Yes" and "No"- Decisions

Each case was subjected separately to an ANOVA, with the number of evaluative statements as dependent variable, in order to find out whether different decisions would be associated with different evaluative patterns across information categories. Statistically, this corresponds to two-or three-way interaction effects that include Decision and Information. Few of the interaction effects were significant. However, comparing the patterns of positive and negative evaluations of those who decided to prescribe with the patterns of those who made the opposite decision could provide some ideas about information strategies. The "Yes" and "No"-groups differed not only in how often they evaluated the central risk factor(s) in a case as favouring drug treatment but also in that the "No"-group seemed to have identified at least one information category as evidence against treatment.

Table 3. Characteristics of the cases and the percentages of doctors who decided to prescribe a drug (Studies III-V). The recommended treatment according to the 1999 Swedish guidelines are also shown.

Information category	Case					
	IS	GM	TW	SH	AR	PU
Age	67	53	67	51	56	41
Sex	Female	Female	Male	Female	Male	Female
Heredity	Slight	Slight	No	Slight	No	Strong
Hypertension	Yes	No	Yes	No	No	No
Diabetes	No	Yes	No	No	No	No
CHD	No	No	No	No	Yes	No
Smoking	No	No	Yes	No	No	No
Overweight	Slight	No	No	No	Yes	No
HDL	1.2	1.1	1.0	1.0	0.9	1.2
LDL	5.4	4.1	3.6	4.3	4.3	5.3
Cholesterol	7.3	5.9	6.0	6.5	5.9	7.2
Decision according to guidelines	No	Yes	Yes	No	Yes	No? *
Percentage of doctors who decided to prescribe	60	50	35	0	85	70

* Heredity is not directly included in the charts but mentioned in a footnote as a factor increasing the coronary risk

Clustering of Participants

A cluster analysis was performed with the purpose of identifying subgroups of doctors with characteristic patterns regarding how the different information

categories were evaluated. Twenty-two variables (the 11 information categories, each with positive and negative directionality) were used. Three almost equally large clusters could be identified. One of them was characterised as having positive evaluations of the cholesterol values as the most frequent category and almost no negative evaluations of cholesterol. A second cluster had almost the opposite pattern of negative and positive evaluations of the cholesterol variable, while a third cluster was characterised by few evaluations for all categories and no distinct pattern.

Disagreement

It was hypothesised that it would be more common for information about lifestyle-related factors like smoking and overweight to be judged differently by different doctors in a particular patient case. Regarding case TW's smoking, there were 21 statements with a positive direction and 10 with a negative direction. For case AR's overweight there were three positive and six negative directions. For the traditional medical risk factors hypertension, diabetes and CHD there was, with minor exceptions, agreement on their evaluation. As could be expected from the cluster analysis, there was also disagreement about how the cholesterol variable was evaluated. For four of the cases there were approximately the same numbers of positive and negative evaluations of the same cholesterol value.

Use of Rules

A total of 32 statements were coded as rules. Of these, 18 were judged to be derived from or compatible with the guidelines (e.g. treating secondary prevention cases). Twenty-four of the 32 rules were expressed in connection with the secondary prevention case (AR) and the diabetes case (GM), and it was hypothesised that the use of rules would be more frequent for participants who decided to prescribe than for the other group. The association was in the expected direction but not significant (chi-square).

Study V

The aim of this study was to investigate how GPs use their knowledge about the guidelines on hyperlipidaemia in their decisions on drug treatment, and what their arguments are for their decisions.

Contents of the Free-Report Protocols

For evaluating the reliability in coding, two of the authors coded 50 randomly selected statements from the Free-Report Protocols independently of each other.

For 78% of the statements there was complete agreement as to the cause variable(s). The same rate of agreement was found for effect variables.

The mean number of relations including the three effect variables Usability, Probability and Implication was 10.0 per doctor. The most frequent effect variable was Probability (mean 6.1), followed by Implication (mean 2.4) and Usability (mean 1.5). Cholesterol Level was the most frequent cause variable, followed by CHD, Diabetes, Family History and Age in decreasing frequency. When the variables were compared with respect to the number of doctors who had included them, Diabetes ranked first followed by Family History.

There were a few relations with cause variables outside of the traditional risk factors. The most common was the Patient's Attitude with Probability as effect for seven participants and Implication for one participant. Other such cause variables were Side Effects of Drugs and Cost-Benefit Considerations.

Evidence of Knowledge Related to Guidelines

The number of doctors who included Diabetes, CHD, Stroke/TIA, and PAD in their Free-Report Protocols and/or Think-Aloud Protocols was analysed. The latter three variables were also analysed as subgroups of CVD, Heart Disease and Cerebrovascular Disease. Of the 20 doctors, 14 indicated that CHD alone, or in combinations with other variables, should lead to drug prescription (Implication). If CVD and Heart Disease were included, the number increased to 17.

For Diabetes, 11 doctors out of 20 had statements with Diabetes as cause and Implication as effect. If Probability was included, practically all the doctors revealed having this knowledge regarding Diabetes and CHD in one of the two protocols.

Few doctors expressed the association regarding Stroke/TIA and PAD even if the more general terms cardiovascular or cerebrovascular disease were included.

Decisions and Arguments

The characteristics of the different cases and their recommended treatment according to the 1999 Swedish guidelines are shown in Table 3.

For the CHD case (AR) the presence of angina pectoris should be a strong indication for drug therapy. Of the 20 doctors, 17 decided to prescribe a drug. All three doctors who refrained from treatment expressed the effect variable Implication in relation to CHD. From the Think-Aloud Protocols we found that one of the three doctors questioned the diagnosis angina pectoris and two

suggested that weight reduction should be further encouraged before starting drug treatment.

For the diabetic case (GM), there was no clear difference between the 10 who decided to prescribe and the 10 with the opposite decision regarding the frequency of statements with Diabetes affecting Probability or Implication for drug prescription. Among the 10 doctors with a "No" decision, three expressed an inclination to prescribe a drug but said at the same time that they wanted to wait and see. Two doctors said that they wanted to await further change in lifestyle. Another three doctors seemed to have identified arguments against treatment (lack of other diseases and risk factors, a mild form of diabetes, marginally elevated cholesterol). Two doctors gave no clear argument for their decisions.

Among the primary prevention cases, TW had the highest 10-year Framingham risk in the range 20-40%, which would justify pharmacological treatment. However, 13 of the 20 doctors decided not to prescribe. As arguments for their decisions, lifestyle factors were common, but both as arguments for prescribing and for not prescribing.

Case PU described a young female with a fairly high cholesterol/LDL and a severe family history of CHD. A simple application of the risk chart would give a mild risk (5-10%) or, more correctly, a moderate risk (10-20%) if the time span were projected to the age of 60 years. However, the severe family history should be taken into account, which would make drug treatment reasonable. All 14 out of 20 who chose to prescribe gave family history as an argument. Of the non-prescribers two seemed to have used the patient's young age as an argument against prescription and one the large number of risk-lowering factors.

Case IS represented a patient with a moderate risk (10-20%), which would ordinarily not motivate drug prescription. Twelve (60%) of the doctors chose to prescribe, and the presence of hypertension was the most frequent argument. Case SH was a low-risk case and all doctors chose not to prescribe. The absence of other risk factors and diseases was most often used as an argument for the decision.

Discussion

The purposes of the studies were to increase our understanding of GPs' decisions on drug treatment of hypercholesterolaemia and to assess the usefulness of different methods for examining decision-making on drug treatment. Studies I, IV and V represent two different methodological approaches (Clinical Judgement Analysis and think-aloud) and they can be compared with respect to the relative importance that different kinds of information seem to have had for the treatment decisions. They can also elucidate in different ways how the information about the patient may have been combined and integrated to reach a decision. Study II concerned the problem with intuitive risk estimates as the basis for distinguishing between high- and low risk patients. Study III aimed at studying the validity of two different response measures for describing the decision process over time, and the results from this study will be discussed in a general methodological context.

Information Affecting the Decisions

In the CJA study the importance of an information category (cue) was defined as its coefficient in the multiple regression equation. A doctor's strategy was described as his or her set of such regression weights. When the decision weights were averaged over the doctors, CHD was the most important variable, followed by cholesterol, while age, diabetes and hypertension were about equal in importance. Among lifestyle-related factors, only smoking was included in the case vignettes (fifteen of the 40 patients were smokers) and it had a statistically significant regression weight for only two of the 38 GPs.

In Studies IV-V the importance of different patient variables was evaluated in two different ways. In Study IV, the number of times a certain variable (information category) was evaluated in a positive or negative direction was a measure of its importance. Cholesterol was the most important, followed by smoking, hypertension, family history and CHD. In Study V a measure of the importance of different patient variables was determined by how often they were included in the cause-effect relations from the Free-Report Protocols. Cholesterol level was the most frequent, followed by CHD, family history, age, and smoking.

The comparison between outcomes from the think-aloud and CJA studies should be interpreted with great caution. First, the CJA study used a more limited set of variables (e.g. family history was not included), and second, the small number of case vignettes in the think-aloud study probably means that the frequency with which different variables are evaluated will be very sensitive to how the small set

of vignettes is constructed. However, one reasonable conclusion is that the degree of cholesterol elevation is important for the prescription decision, regardless of the method used. Both methods indicated that CHD, hypertension and diabetes were important contributors to the decision to prescribe or not, and their relative importance could be method dependent. The only lifestyle-related factor that was present in both CJA and think-aloud was smoking, and it seemed to be more important in the coding of verbal protocols than in the assigning of statistical regression weights. Family history, which is not directly included in risk charts, was among the most frequent categories in the verbal protocols (but it was not included in the CJA study).

Regarding individual differences, Study I was designed so that they could be described as the patterns of regression weights. As in other CJA studies on medical decision-making (Wigton 1996), the results indicated a great diversity in strategies. The differences between doctors were also evident in terms of how consistent they were in their judgements (explained variance, or r^2) according to the linear multiple regression model used. In the cluster analysis (Study IV) the most important factor that discriminated the participants was the extent to which they emphasised the degree of cholesterol elevation, and the extent to which they used low cholesterol values (i.e. close to normal values) as speaking against treatment.

If the criteria for importance are widened to include factors in addition to statistical weights and frequency of occurrence in the verbal protocols, it can be argued that lifestyle factors seemed to be important in the sense that they often caused ambiguity or disagreement, and were often used as arguments for not following the guidelines strictly.

Evaluation and Integration of Information

The CJA study does not provide any knowledge about how the information is integrated into a decision apart from telling us how close the judgements could be fitted to the linear regression model. In other words, the model presupposes that a given value of a variable (e.g. a certain level of cholesterol elevation or female sex) has the same influence regardless of the values of other variables. On the other hand, it is easy to find examples of when such interactions between variables are reasonable. For instance, the degree of cholesterol elevation should probably be of less importance for secondary prevention cases than for primary prevention. One conclusion can be that CJA is rather limited when studying the decision area, at least in relation to the guidelines currently in use.

The think-aloud study gives more information and hypotheses about how the decision could be characterised than is possible with CJA. Study III tested whether the decision process was compatible with two theories on how decision alternatives are evaluated over time: the Search for Dominance Structure Theory (Montgomery 1983, 1989), and the Differentiation and Consolidation Theory (Svenson 1992, 1996). Both theories predict that the evaluative difference between a preliminarily chosen alternative and its competitor (in the present case between the decisions to prescribe and not to prescribe) will increase over time. The data supported this prediction. In the same way that preliminary diagnostic hypotheses are generated early in the patient encounter (Elstein 1978), this may be also be true for early preliminary therapeutic decisions.

In Study IV the “Yes” groups and “No” groups were compared with regard to their patterns of evaluations of different information categories. Although the statistical evidence was weak, there were indications that non-prescribers tended to evaluate central risk factors with a positive directionality less often than prescribers. They also appeared to identify at least one information category that was given a negative directionality. This is compatible with theories that describe decision-making as a search for arguments or reasons for one or the other decision alternative (Montgomery 1983, Lipshitz 1993, Shafir 1993). The information categories that some doctors used as arguments against treatment were used as arguments for treatment by other doctors. The findings can be interpreted as showing that doctors evaluate given information from different perspectives, i.e., from different viewing angles that will put different aspects of the information in the foreground and background, respectively (Montgomery 1994). If we take smoking or overweight as examples, they could be seen as risk factors for CHD, which would increase the inclination for drug prescription, or as possibilities for lifestyle change and as arguments for refraining from or postponing drug treatment.

Such disagreement in the evaluation of given information about a patient was demonstrated not only for overweight and smoking but also for age and cholesterol. With these variables the disagreement can be discussed in terms of different cut-off values rather than different perspectives. Age is generally considered as positively and monotonically related to risk for CHD. At the same time, the guidelines make the reservation that the benefit of giving drugs to very old people is unclear. It can be noted that in the CJA study age had a negative average weight, which means that an age of 65 or above was associated with a lower probability for prescription than an age below 65 (the age variable was dichotomised). As far as young patients are concerned, the perspective of 10-year

risk appears to be too narrow. The recommended procedure is to project the age to 60 years in order to estimate the risk (Läkemedelsverket 1999, Wood 1998). For doctors with limited experience in using the risk charts, this might be confusing.

One and the same cholesterol value for a certain patient was valued as favouring drug treatment for some doctors and disfavouring treatment for others. The selection of cases could be part of the explanation for this disagreement. Three of the six cases had cholesterol values in the range of 5.0-6.5mmol/l, which is often labelled as mild elevation. This might have formed the basis for negative evaluations, i.e. when a value was close to normal a decision to refrain from treatment could have been favoured.

The issue of seeing drug treatment as a search for arguments and the application of different perspectives was further developed in Study V by searching the protocols for arguments for the different decisions.

Decisions and Risk Estimates Compared with Guidelines

In the CJA study the doctors were not asked for a Yes/No decision but merely to state their inclination for drug treatment on a graded scale. For 17 of the cases there was no support for pharmacological intervention according to the guidelines. All of these cases had a median judgement below or close to 50% (prescribing and non-prescribing equally reasonable). For the 12 secondary prevention cases all but two had a median rating above 50%. In other words the ratings seemed to be reasonable on a group level.

In the think-aloud study 17 of the 20 doctors chose prescription for the secondary prevention case. The three doctors who chose not to prescribe seemed to be aware of the message from the guidelines based on their verbal protocols, and they gave explicit arguments for their decisions. The same conclusion could be drawn concerning the diabetes case: the doctors exhibited good knowledge of guidelines, and the decisions to refrain from treatment were reasonable when considering the other characteristics of the case, e.g. a cholesterol value close to normal. The decision to prescribe or not was included in Study II, where the primary goal was to investigate the risk estimates. Two secondary prevention cases were included, and the percentage of doctors who were inclined to treat these was 39% and 49%, respectively. Part of the explanation for the discrepancy compared to the 85% willing to treat the secondary prevention case in the think-aloud study can be methodological differences. On the other hand, Study II was conducted just before the 1999 Swedish guidelines were published, with their emphasis on

secondary prevention. Put another way, the data indicate that the 1999 guidelines concerning the role of CHD have had a good impact.

The risk estimates for both primary and secondary prevention showed tendencies towards underestimation of risk. The results are somewhat contradictory compared to a number of other studies on doctors' risk estimates based on simulated patients (Grover 1995, Friedman 1996, Pignone 2003). The reason for this is not clear. The clearest finding in Study II was underestimation for high-risk patients (over a 20% 10-year risk). The only cholesterol case included in the Friedman study was a moderate risk case (10-20% 10-year risk when the Framingham risk chart is used). In the studies by Pignone et al and Grover et al a wide range of risk was present. It should be noted that the two patient cases in the Pignone study with the highest risks (both with a 20-40% 10-year risk according to charts, and a 22% and 18% 5-year risk, respectively, according to the article) both had accurate median risk estimates (25% and 20%, respectively). Another issue of interest, about which we have little information, is how risk estimates correspond to treatment decisions on simulated or real patients. In Study II there was a poor correspondence between the proportion of doctors who wanted to treat a case and the corresponding median risk estimate (Spearman .46, ns).

Methodological Considerations

Participating Doctors

The conclusions from the studies, and in particular the comparisons between studies should be regarded with caution because of the sizes of the non-responding groups. In particular, the response rate was low in Study II. The population of doctors also differed between the studies. Study II targeted a random sample from all Swedish GPs, while Studies I and III-V included GPs from the southern Stockholm area, and the selection of doctors in the think-aloud study was not randomised. There are probably differences between different areas in Sweden, e.g. with regard to how guidelines have been implemented, which could have affected the drug prescription decisions. We cannot be totally sure that the selection of study groups and the dropout frequencies have not affected the results regarding information processing strategies, ability to make risk estimates, and so on.

Validity of Case Vignettes

The importance of the results rests on whether case vignettes lead to the same decisions and actions as for real patients, i.e. their validity (or criterion validity). Case vignettes or written case descriptions have provided a convenient way to

study medical decisions and judgements. Well-defined information can be presented to different doctors and the information content can be varied systematically. However, little is known about their criterion validity.

A systematic review of the evidence concerning the validity was carried out by Jones and co-workers (1990). They selected articles in which a measure of judgement or behaviour in actual clinical practice was compared with the doctors' responses to case vignettes. Eleven articles that met this criterion were assessed. According to the authors, only two of these were designed so that a comparison between clinical behaviour and behaviour with the simulations could be assessed. Both of the studies dealt with judgements concerning rheumatic activity in patients with rheumatoid arthritis. In the first study (Kirwan 1983 a) the doctors assessed a number of patients regarding disease activity, and several weeks later they judged a number of case vignettes, including some constructed from their own patients. There was a strong correlation between the measures for the real patients and the corresponding case vignettes ($r = .90$). The study was replicated (Kirwan 1983 b) with another group of rheumatologists and gave similar results ($r = .85$). The findings support the idea that under some circumstances case vignettes can have a high degree of validity. However, the possibility that the doctors remembered their patients and the judgements they made then has not been assessed.

A more recent study with a comparison between the actions taken with case vignettes and the actions with corresponding matched patients was conducted by Sandvik (1995) and concerned women with urinary tract infections. A higher activity level on behalf of the doctors seemed to be associated with vignettes as compared with real patients, at least when response options were given, a so-called cueing effect.

In the present studies the criterion validity was not evaluated. However, the 20 participants in the think-aloud study were questioned about the realism of the cases and the mode of presentation, and almost everyone found the scenarios realistic, which supports the face validity of the setting. As far as the treatment decisions, and thus criterion validity, are concerned, it can be suspected that the doctors tried to follow what they believed to be a correct decision, thereby performing "better" with the vignettes than with real patients. Study III aimed at examining the validity of rating scales and verbal protocols. The criterion for comparison was whether the two measures would reflect the changes in decision directionality over time in a similar way, as these changes were predicted by two process theories of decision-making. The two measures both followed the predictions, which is taken as support of their validity in a theoretical sense (construct validity, Helmstadter 1964).

The reliability and validity of the present studies are dependent on the response scales and coding procedures used. In Study I, and for half the participants in study III, a VAS scale was used. The ratings were analysed with linear regression and ANOVA, which requires an interval scale. In a strict sense the ratings could be considered as being on an ordinal level (Campbell 1998), but the instructions to rate in percentages was supposed to make the scale close to an interval scale. In study III, there was evidence of a “floor” effect of the rating scale. Participants who rated low on the scale on the first screens had little room to lower their ratings further. The rating scale may also have influenced some participants to be committed to a certain decision; if their ratings were above the midpoint at the end of the series of screens there may have been pressure to be consistent and answer “Yes” to the final question as to whether they would prescribe or not.

The coding of the verbal protocols proved to have a fairly good inter-judge reliability. The criteria for whether directionality should be assigned or not (studies III-IV) or whether the effect category qualified as Probability or Implication were not self-evident, which may have affected some of the results. The grading of directionality into “+” and “-” may have been too crude to detect leaps in the inclination towards a certain decision, which could be better reflected by the rating scale. Finally, the risk estimates (Study II) were made by asking for the estimated number in a group of 100 hypothetical individuals with identical risk factors. It was supposed that this would facilitate risk judgements (Gigerenzer 1996). On the other hand, the majority of other studies with the same interest have used risk levels in percentages, which may make comparison with these other studies more difficult.

Evaluations of the Decisions

Evaluation of the doctors’ decisions was done by comparing them to the guidelines. The guidelines refer to the Framingham equation for determining the CHD risks for primary prevention cases. If the GPs’ decisions on these cases are compared to the Framingham formula, the latter should be valid for the Swedish population. Several recent papers have raised concern about the generalisability of the Framingham equation. Hense (2003) demonstrated that Framingham risks substantially overestimated the actual risks observed in two German cohorts. The conclusion was that local guidelines need to correct for this overestimation. De Visser (2003) applied the Copenhagen Risk Score (Thomsen 2001 b) to a Dutch population. The correlation between the resulting sets of risks was high (.94) but the Copenhagen Risk Score resulted in lower risks than the Framingham equation. The difference between the two estimates became larger with increasing risk.

Bastuji-Garin (2002) evaluated the accuracy of the Framingham risk equation in predicting CHD in seven European populations. The predicted CHD rates were at least twice as high as the observed rates, and for the pooled Scandinavian populations the ratio was 2.8.

A reasonable conclusion is that the Framingham equation probably overestimates coronary risks in patients seen by Swedish GPs. If this equation is applied, it may have consequences for health-economic calculations. The consequences for clinical practice depend on the validity of the cut-off value chosen in the guidelines (e.g. 20% within 10 years). Large differences in the distribution of risk factors between different populations (which was shown in the Bastuji-Garin 2002 study) may also be relevant for the validity of prediction equations.

Conclusions

Guidelines stress the importance of treating patients at highest risk, in particular those with already established cardiovascular disease or diabetes. On the whole, the decisions and knowledge of the GPs in this respect seemed to be congruent with the guidelines.

Compared with the predictions from the Framingham equation, GPs and students showed a tendency to underestimate patients' risks for future coronary disease. There is a need to improve the availability and usability of guidelines, and to update them with relevant risk data and risk assessment tools.

GPs' strategies in terms of the relative weights given to different patient variables varied widely between doctors. The doctors seemed to have different opinions about how to evaluate the degree of cholesterol elevation and how to evaluate lifestyle factors. Smoking and overweight could be regarded as either contributing to the total cardiovascular risk and thereby increasing the inclination for drug prescription, or as representing alternative choices for action rather than drug prescription. The role of lifestyle factors in the decisions on drug treatment could possibly be clarified in future guidelines on cardiovascular prevention.

CJA studies with resulting regression weights and think-aloud studies with coded verbal protocols give different and complementary information about doctors' decisions. The analyses of data in multiple regression research is straightforward, individual differences can be described and quantified, and it is also possible to explore the relation between the doctors' actual strategies and their beliefs about their strategies. Think-aloud data give richer information about the doctors' thinking, but the coding of data is time-consuming. The coding scheme for think-aloud data that was developed in this thesis proved to have good reliability and validity for describing how a treatment decision changes over time and should be a valuable tool for future research.

Forthcoming Research

The think-aloud technique with coding protocols for aspects such as decision directionality and decision arguments should be appropriate in research with relevance for medical teaching. A comparison between experienced clinicians and medical students could give information about aspects of expertise that will both increase our theoretical knowledge and be useful in teaching.

The integration of decision supports into the computerised medical record should benefit from knowledge about how difficult therapeutic and diagnostic decisions could be described in cognitive psychological terms. Examples from this thesis are the difficulties in risk estimates without calculation tools, and the ambiguity of some types of information, such as lifestyle factors, as arguments for drug prescription. A first step could be to use the presentation technique in Studies III-V and study how various formats of a risk calculation tools are used with regard to feasibility and effects on decisions.

The use of drugs for reducing cardiovascular risk factors such as hypercholesterolaemia and hypertension should rest on risk estimates. Different populations have a large spread in overall risk, and the relative importance of different risk factors may also differ from one country to another. A study of risk assessments and treatment decisions for patient cases made by doctors from a high risk country and a low risk country should be of interest regarding further implementation of guidelines.

Acknowledgements

A number of people have provided indispensable help during my work on this thesis. In particular, I wish to express my sincere gratitude to:

Associate Professor *Lars-Erik Strender*, my supervisor, for constant interest, support and for many wise ideas about how to proceed with the project.

Professor *Henry Montgomery*, my associate supervisor, for guiding me back to cognitive psychology, for sharing outstanding knowledge, and for always being optimistic and supportive.

Professor *Jan Sundqvist* and Professor Emeritus *Hans Åberg*, for giving me the opportunity to do my thesis work at Family Medicine Stockholm.

Associate Professor *Johan Bring*, for statistical advice and for formulating many crucial questions.

Ylva Skånér, for collaboration and for many fruitful discussions.

Professor *Ola Svenson*, Department of Psychology, University of Stockholm, for invaluable comments on the project.

Kersti Ejeby, my primary-health-care employer and colleague, for stressing the importance of research in primary health care and for giving me the opportunity to do this research.

Birgitta Danielsson, for collaboration and for much good advice.

Linnea Löwgren, for transcribing the protocols.

Jane Wigertz, for revising the English text.

Wilhelm Montgomery, for help with the software QA™ for computer presentation of patient cases

The whole staff and fellow research students at Family Medicine Stockholm, for their interest and for fruitful discussions.

My fellow GPs and the rest of the staff at Gustavsberg Health Centre, for providing an inspiring arena for primary- health- care work.

The GPs and students who participated in the studies

My family – *Lena, Tobias, David, Fanny* and *Filippa* – for love and support.

This thesis was supported by grants from the Stockholm County Council and the Swedish Heart Lung Association.

References

Abookire S.A., Karson A.S., Fiskio J. and Bates D.W. (2001) Use and monitoring of "statin" lipid-lowering drugs compared with guidelines. *Arch Intern Med* 161, 2626-2627.

[The] AFCAPS/TexCAPS Research Group (1998) Primary prevention of acute coronary events with lovastatin in men and women with average cholesterol levels. Results of AFCAPS/TexCAPS. *JAMA* 279, 1615-1622.

Allen V.A., Arocha J.F. and Patel V.L. (1998) Evaluating evidence against diagnostic hypotheses in clinical decision making by students, residents and physicians. *Med Inf (Lond)* 51, 91-105.

Anderson K.M., Wilson P.W.F., Odell P.M. and Kannel W.B. (1991) An updated coronary risk profile. A statement for health professionals. *Circulation* 83, 356-362.

Anderson K.M., Odell P.M., Wilson P.W.F. and Kannel W.B. (1990) Cardiovascular disease risk profiles. *Am Heart J* 121, 293-298.

Arocha J.F., Patel V.L. and Patel Y.C. (1993) Hypothesis generation and the coordination of theory and evidence in novice diagnostic reasoning. *Med Dec Making* 13, 198-211.

Axelrod R. (1976) The analysis of cognitive maps. In *Structure of Decision*. Edited by Axelrod R. Princeton Univ Press, Princeton. pp 55-73.

Avis N.E., Smith K.W. and McKinlay J.B. (1989) Accuracy of perceptions of heart attack risk: what influences perceptions and can they be changed? *Am J Public Health* 79, 1608-1612.

Bakx J.C. (2000) Changes in serum total cholesterol levels over 18 years in a cohort of men and women: The Nijmegen Cohort Study. *Prev Med* 30, 138-145.

Bastuji-Garin S., Deverly A., Moyses D., Castaigne A., Mancia G., de Leeuw P.W., Ruilope L.M. and Chatelier G. (2002) The Framingham prediction rule is not valid in a European population of treated hypertensive patients. *J Hypertens* 20, 1937-1938.

Berglund G., Eriksson K.F., Israelsson B., Kjellström T., Lindgärde F., Mattiasson I., Nilsson J.Å. and Stavenow L. (1996) Cardiovascular risk groups and mortality in an urban Swedish male population: the Malmö preventive project. *J Intern Med* 239, 489-497.

GPs' decision-making on drug treatment of hypercholesterolaemia

Bjerrum L., Larsen J. and Kragstrup J. (2001) Guidelines accompanied by changes in reimbursement. *Scand J Prim Health Care* 19, 158-162.

Brehmer B. (1987) The psychology of risk. In *Risk and decision*. Edited by Sigleton W.T. and Hovden J. John Wiley & Sons Ltd. pp 25-39.

Brehmer A. and Brehmer B. (1988) What have we learned about human judgment from thirty years of policy capturing? In *Human judgment: the SJT view*. Edited by Brehmer B. and Joyce C.R.B. North Holland/Elsevier Science Publisher, New York. pp 75-114.

British Cardiac Society, British Hyperlipidaemia Association, British Hypertension Society, endorsed by the British Diabetic Association (1998) Joint British recommendations on prevention of coronary heart disease in clinical practice. *Heart* 80 (supplement 2) S1-S29.

Brownstein A. L. (2003) Biased predecision processing. *Psychol Bull* 129, 545-568.

Bucher H.C., Weinbacher M. and Gyr K. (1994) Influence of method of reporting study results on decision to prescribe drugs to lower cholesterol concentration. *BMJ* 309, 761-764.

Burgers J., Grol R., Zaat J., Spies T., van der Bij A. and Mokkink H. (2003) Characteristics of effective clinical guidelines for general practice. *Br J Gen Pract* 53, 15-19.

Cabana M.D., Rand C.S., Powe N.R., Wu A.W., Wilson M.H., Abboud P.-A.C. and Rubin H.R. (1999) Why don't physicians follow clinical practice guidelines? A framework for improvement. *JAMA* 282, 1458-1465.

Campbell W.I., and Patterson C.C. (1998) Quantifying meaningful change in pain. *Anaesthesia* 53, 121-125.

Charlin B., Tardiff J. and Boshuizen H.P.A. (2000) Scripts and medical diagnostic knowledge: Theory and applications for clinical reasoning and instruction research. *Acad Med* 75, 182-190.

[The] Cholesterol and Recurrent Events Trial Investigators (1996) The effect of Pravastatin on coronary events after myocardial infarction in patients with average cholesterol levels. *N Engl J Med* 335, 1001-1009.

Cook R.J. and Sackett D.L. (1995) The number needed to treat: a clinically useful measure of treatment effect. *BMJ* 310, 452-454.

- Cooksey R.W. (1966) *Judgment analysis: theory, methods, and applications*. Boston: Academic Press.
- Cranney M. and Walley T. (1996) Same information, different decisions: the influence of evidence on the management of hypertension in the elderly. *Brit J Gen Practice* 46, 661-663.
- Crozier W.R. and Ranyard R. (1997) Cognitive process models and explanations of decision making. In *Decision making*. Edited by Ranyard R., Crozier W.R. and Svenson O. Routledge, London. pp 5-33.
- Custers E.J.F.M., Boshuizen H.P.A. and Schmidt H.G. (1996) The influence of medical expertise, case typicality, and illness script component on case processing and disease probability estimates. *Mem Cognit* 24, 384-399.
- Custers E.J.F.M., Boshuizen H.P.A. and Schmidt H.G. (1998) The role of illness scripts in the development of medical diagnostic expertise: results from an interview study. *Cogn Instr* 16, 367-398.
- Dahlstrand U. and Montgomery H.M. (1984) Information search and evaluative processes in decision making: a computer-based process-tracing study. *Acta Psychol (Amst)* 56, 113-123.
- Dawber T., Meadors G. and Moore F. (1951) Epidemiological approaches to heart disease: The Framingham study. *Amer J Publ Health* 41, 279-286.
- Dawson N.V. and Arkes H.R. (1987) Systematic errors in medical decision making: Judgment limitations. *J Gen Intern Med* 2, 183-187.
- Dhami M.K. and Harries C. (2001) Fast and frugal versus regression models of human judgement. *Thinking and Reasoning* 7, 5-27.
- Doherty M. and Brehmer B. (1997) The paramorphic representation of clinical judgment: A thirty-year retrospective. In *Research on judgment and decision making: Currents, connections, and controversies*. Edited by Goldstein, W.M. and Hogarth R.M. Cambridge series on judgment and decision making. pp 537-551.
- Edwards A., Elwyn G. and Mulley A. (2002) Explaining risks: turning numerical data into meaningful pictures. *BMJ* 324, 827-830.
- Elstein A.S., Schulman L.S. and Sprafka S.A. (1978) *Medical problem solving: An analysis of clinical reasoning*. Harvard University Press, Cambridge, MA.
- Ericsson K.A. and Simon H.A. (1980) Verbal protocols as data. *Psychol Rev* 87, 215-251.

GPs' decision-making on drug treatment of hypercholesterolaemia

Ericsson K.A. and Simon H.A. (1999) *Protocol Analysis: Verbal reports as data*. Revised version The MIT Press, Cambridge, MA, US.

EUROASPIRE Study Group (1997) EUROASPIRE. A European Society of Cardiology survey of secondary prevention of coronary heart disease: Principal results. *Eur Heart J* 18,

EUROASPIRE I and II Group (2001) Clinical reality of coronary prevention guidelines: a comparison of EUROASPIRE I and II in nine countries. *Lancet* 357, 995-1000.

Evans J.S., Harries C., Dennis I. and Dean J. (1995) General practitioners' tacit and stated policies in the prescription of lipid lowering agents. *Br J Gen Pract* 45, 15-18.

Evans A., Tolonen H., Hense H.W., Ferrario M., Sans S. and Kuulasmaa K. (2001) Trends in coronary risk factors in the WHO Monica project. *Int J Epidemiol* 30 (supplement 1), S35-S40.

Field M.J. and Lohr M.J. eds (1990) *Clinical practice guidelines: directions for a new program*. National Academic Press, Washington, DC.

Fisch H.-U., Hammond K.R., Joyce C.R.B. and O'Reilly M. (1981) An experimental study of the clinical judgment of general physicians in evaluating and prescribing for depression. *Br J Psychiatry* 138, 100-109.

Friedman P.D., Brett A.S. and Mayo-Smith M.F. (1996) Differences in generalists' and cardiologists' perceptions of cardiovascular risk and the outcomes of preventive therapy in cardiovascular disease. *Ann Intern Med* 124, 414-421.

Gigerenzer G. (1996) The psychology of good judgment: frequency formats and simple algorithms. *Med Decis Making* 16, 273-280.

Greenland P., Knoll M.D., Stamler J., Neaton J.D., Dyer A.R., Garside D.B. and Wilson P.W. (2003) Major risk factors as antecedents of fatal and nonfatal coronary heart disease events. *JAMA* 289, 891-897.

Grimshaw J.M. and Russell I.T. (1993) Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. *Lancet* 341, 1317-1322.

Grover S.A., Lowensteyn I., Esrey K.I., Steinert Y., Joseph L. and Abrahamowicz M. (1995) Do doctors accurately assess coronary risk in their patients? Preliminary results of the coronary health assessment study. *BMJ* 310, 975-978.

Harvald B. (1996) Making doctors change. *J Intern Med* 239, 191-193.

References

- Heart Protection Study Collaborative Group (2002) MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20 536 high-risk individuals: a randomised placebo-controlled trial. *The Lancet* 360, 7-22.
- Helmstadter G C. (1964) Principles of psychological measurement. New York: Appleton-Century-Crofts.
- Hense H.W., Schulte H., Lowel H., Assmann G. and Keil U. (2003) Framingham risk function overestimates risk of coronary heart disease in men and women from Germany– results from the MONICA Augsburg and the PROCAM cohorts. *Eur Heart J* 24, 937-945.
- Hobbs F.D. and Erhardt L. (2002) Acceptance of guideline recommendations and perceived implementation of coronary heart disease prevention among primary care physicians in five European countries: the Reassessing European Attitudes about Cardiovascular Treatment (REACT) survey. *Fam Pract* 19, 596-604.
- Johannesson M. (2001) At what coronary risk level is it cost-effective to initiate cholesterol lowering treatment in primary prevention? *Eur Heart J* 22, 919-925.
- Johannesson M., Jönsson B., Kjekshus J., Olsson A.G., Pedersen T.R. and Wedel H. (1997) Cost effectiveness of simvastatin treatment to lower cholesterol levels in patients with coronary heart disease. *N Engl J Med* 336; 332-336.
- Jones T.V., Gerrity M.S. and Earp J. (1990) Written case simulations: do they predict physicians' behavior? *J Clin Epidemiol* 43, 805-815.
- Jones A.F., Walker J., Jewkes C., Game F.L., Bartlett W.A., Marshall T. and Bayly G.R. (2001) Comparative accuracy of cardiovascular risk prediction methods in primary care patients. *Heart*. 85, 37-43.
- Joseph G.-M. and Patel V.L. (1990) Domain knowledge and hypothesis generation in diagnostic reasoning. *Med Decis Making* 10, 21-46.
- Kirwan J.R., Chaput de Saintonge D.M., Joyce C.R.B. and Currey H.L.F. (1983 a) Clinical judgment in rheumatoid arthritis. I. Rheumatologists' opinions and the development of 'paper patients'. *Ann Rheum Dis* 42, 644-647.
- Kirwan, J.R., Bellamy N., Condon H., Buchanan W.W. and Barnes C.G. (1983 b) Judging "current disease activity" in rheumatoid arthritis-an international comparison. *J Rheumatol* 10, 901-905.
- Kirwan J.R., Chaput de Saintonge D.M. and Joyce C.R.B. (1990) Clinical judgment analysis. *Q J Med* 281, 935-949.

- Klayman J. (1988) On the how and why (not) of learning from outcomes. In Human judgment: the SJT view. Edited by Brehmer B. and Joyce C.R.B. North Holland/Elsevier Science Publisher, New York. pp 75-114.
- Knuiman M.W., Vu H.T.V. and Segal M.R. (1997) An empirical comparison of multivariate methods for estimating risk of death from coronary heart disease. *J Cardiovasc Risk* 4, 127-134.
- Kushniruk A.W., Patel V.L. and Marley A.A.J. (1998) Small worlds and medical expertise: implications for medical cognition and knowledge engineering. *Int J Med Inf* 49, 255-271.
- Kushniruk A., Patel V. and Fleischer D. (1995) Analysis of medical decision making: a cognitive perspective on medical informatics. *Proc Annu Symp Comput Appl Med Care*, 193-197.
- Lichtenstein S., Slovic P., Fischhoff B., Layman M. and Combs B. (1978) Judged frequency of lethal events. *J Exp Psychol [Hum Learn]* 4, 551-578.
- Lipshitz R. (1993) Decision making as argument-driven action. In *Decision making in action: Models and methods*. Edited by Klein G.A., Orasanu J. and Calderwood R. Ablex, Norwood, NJ. pp 172-181.
- Lomas J., Anderson G.M., Domnick-Pierre K., Vayda E., Enkin M.W. and Hannah W.J. (1989) Do practice guidelines guide practice? The effects of a consensus statement on the practice of physicians. *N Engl J Med* 321, 1306-1311.
- [The] Long-term Intervention with Pravastatin in ischaemic disease (1998) Prevention of cardiovascular events and death with Pravastatin in patients with coronary heart disease and a broad range of initial cholesterol levels. *N Engl J Med* 339, 1349-1357.
- Läkemedelsverket. (1995) Behandling av hyperlipidemi. Rekommendationer [Treatment of hyperlipidaemia. Recommendations, in Swedish]. *Information från Läkemedelsverket* 3, 178-203.
- Läkemedelsverket.(1999) Behandling av hyperlipidemi. Rekommendationer. [Treatment of hyperlipidaemia. Recommendations, in Swedish]. *Information från Läkemedelsverket* 7, 15-21.
- Läkemedelsverket.(2003) Behandling av hyperlipidemi- nationella och internationella guidelines [Treatment of hyperlipidaemia- national and international guidelines, in Swedish]. *Information från Läkemedelsverket* 4, 21-68.
- Mansfield C.D. (1995) Attitudes and behaviours towards clinical guidelines: the clinicians' perspective. *Qual Health Care* 4, 250-255.

References

- Meland E., Laerum E. and Stensvold I. (1994) Assessment of coronary heart disease risk, I. A postal inquiry among primary care physicians. *Fam Pract* 11, 117-121.
- Mellers B.A., Schwartz A. and Cooke A.D.J. (1998) Judgment and decision making. *Annu Rev Psychol* 49, 447-477.
- Montgomery H. (1983) Decision rules and a search for a dominance structure: towards a process model of decision making. In *Analyzing and aiding decision processes*. Edited by Humphreys P., Svenson O. and Vari A. North-Holland and Hungarian Academic Press, Amsterdam/Budapest. pp 343-369.
- Montgomery H. (1989) From cognition to action: The search for dominance in decision making. In *Process and Structure in Human Decision Making*. Edited by Montgomery H. and Svenson O. John Wiley & Sons Ltd, Chichester, New York. pp 23-49.
- Montgomery H. (1994) Towards a perspective theory of decision making and judgment. *Acta Psychol (Amst)* 87, 155-178.
- Murray C.J. and Lopez A.D. (1997) Mortality by cause for eight regions of the world: Global Burden of Disease Study. *The Lancet* 349, 1269-1276.
- Niknian M., McKinlay S.M., Rakowski W. and Carleton R.A. (1989) A comparison of perceived and objective CVD risk in a general population. *Am J Public Health* 79, 1653-1654.
- Nisbett R.E. and Wilson T.D. (1977) Telling more than we can know: verbal reports on mental processes. *Psychol Rev* 84, 231-259.
- Patel V.L. and Groen G.J. (1986) Knowledge based solution strategies in medical reasoning. *Cognit Sci* 10, 91-116.
- Pignone M., Phillips C.J., Elasy T.A. and Fernandez A. (2003) Physicians' ability to predict the risk of coronary heart disease. *BMC Health Serv Res*. Jul 11;3(1):13.
- Poses R.M., Cebul R.D., Collins M. and Fager S.S. (1985) The accuracy of experienced physicians' probability estimates for patients with sore throats. *JAMA* 254, 925-929.
- Prosser L.A., Stinnett A.A., Goldman P.A., Williams L.W., Hunik M.G., Goldman L. and Weinstein M.C. (2000) Cost-effectiveness of cholesterol-lowering therapies according to selected patient characteristics. *Ann Intern Med* 132, 769-779.

- Ramsay L.E., Haq I.U., Jackson P.R., Yeo W.W., Picjkin D.M. and Payne, J.N. (1996) Targeting lipid-lowering drug therapy for primary prevention of coronary disease: an updated Sheffield table. *Lancet* 348, 387-388.
- Redelmeier D.A., Koehler D.J., Liberman V. and Tversky A. (1995) Probability judgment in medicine: discounting unspecified possibilities. *Med Decis Making* 15, 227-230.
- Sandvik H. (1995) Criterion validity of responses to patient vignettes: An analysis based on management of female urinary incontinence. *Fam Med* 27, 388-392.
- Scandinavian Simvastatin Survival Study Group. (1994) Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study (4S). *Lancet* 344, 1383-1389.
- Schmidt H.G. and Boshuizen H.P.A. (1993) On acquiring expertise in medicine. *Educ Psychol Rev* 5, 205-221.
- Schmidt H.G., Norman C.R. and Boshuizen H.P.A. (1990) A cognitive perspective on medical expertise: theory and implications. *Acad Med* 65, 611-621.
- Schooler J.W., Ohlsson, S and Brooks, K (1993) Thoughts beyond words: When language overshadows insight. *J Exp Psychol Gen* 122, 166-183.
- Schulman K.A., Escarge J.J., Eisenberg J.M., Hershey J.C., Young M.J., McCarthy D.M. and Williams S.V. (1992) Assessing physicians' estimates of the probability of coronary artery disease. *Med Decis Making* 12, 109-114.
- Shafir E., Simonson I. and Tversky A. (1993) Reason-based choice. *Cognition* 49, 11-36.
- Sjöberg L. (1998) Worry and risk perception. *Risk Anal* 18, 85-93.
- Skånér Y., Strender L.-E. and Bring J. (1998) How do GPs use clinical information in their judgements of heart failure? *Scand J Prim Health Care* 16, 95-100.
- Skånér Y., Bring J., Ullman B. and Strender L.-E. (2000) The use of clinical information in diagnosing chronic heart failure: A comparison between general practitioners, cardiologists and students. *J Clin Epidemiol* 53, 1081-1088.
- Slovic P. (1987) Perception of risk. *Science* 236, 280-285.
- Smith S. J., Blair S. N., Bonow R.O., Brass L.M., Cerqueira M.D. et al. (2001) AHA/ACC Scientific Statement: AHA/ACC guidelines for preventing heart attack and death in patients with atherosclerotic cardiovascular disease: 2001 update: a statement for healthcare professionals from the American Heart Association and the American College of Cardiology. *Circulation* 104, 1577-1579.

References

- Socialstyrelsens läkemedelsavdelning. (1988) Behandling av hyperlipidemi. [Treatment of hyperlipidemia, in Swedish]. Information från Socialstyrelsens läkemedelsavdelning, 152-155.
- Svenson O. (1979) Process descriptions of decision making. *Organ Behav Hum Perform* 23, 86-112.
- Svenson O. (1992) Differentiation and Consolidation Theory of human decision making: A frame of reference for the study of pre- and post-decision processes. *Acta Psychol (Amst)* 80, 143-168.
- Svenson O. (1996) Decision making and the search for fundamental psychological regularities: What can be learned from a process perspective? *Org Behav Hum Decis Process* 65, 252-267.
- Tanner W.P. and Swets J.A. (1954) A decision-making theory of visual detection. *Psychol Rev* 61, 401-409.
- Tape T.G. and Wigton R.S. (1989) Medical students' and residents' estimate of cardiac risk. *Med Dec Making* 9, 170-175.
- Third Joint Task Force of European and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (2003) European guidelines on cardiovascular disease prevention in clinical practice. *Eur Heart J* 24, 1601-1610.
- Thomsen T.F., Jorgensen T., Ibsen H. and Borch-Johnsen K. (2001 a) Assessment of coronary risk in general practice in relation to the use of guidelines: a survey in Denmark. *Prev Med* 33, 300-304
- Thomsen T.F., Davidsen M., Ibsen H., Jorgensen T., Jensen G. and Borch-Johnsen K. (2001 b) A new method for CHD prediction and prevention based on regional risk scores and randomized clinical trials; PRECARD and the Copenhagen Risk Score. *J Cardiovasc Risk* 8, 291-297.
- Truelsen T., Mahonen M., Tolonen H., Asplund K., Bonita R. and Vanuzzo D. (2003) Trends in stroke and coronary heart disease in the WHO MONICA Project. *Stroke* 34, 1346-1352.
- Tversky A. and Kahneman D. (1974) Judgement under uncertainty: Heuristics and biases. *Science* 85, 1124-1131.
- Watkins C., Harvey I., Langley C., Gray S. and Faulkner A. (1999) General practitioners' use of guidelines in the consultation and their attitudes to them. *Brit J Gen Practice* 49, 11-15.

GPs' decision-making on drug treatment of hypercholesterolaemia

[The] West of Scotland Coronary Prevention Study Group. (1995) Prevention of coronary heart disease with pravastatin in men with hypercholesterolaemia. *N Engl J Med* 16, 1301-1306.

Wigton R.S. (1996) Social judgement theory and medical judgement. *Thinking and reasoning* 2, 175-190.

de Visser C.L., Bilo H.J., Thomsen T.F., Groenier K.H. and Meyboom-de Jong B. (2003) Prediction of coronary heart disease: a comparison between the Copenhagen risk score and the Framingham risk score applied to a Dutch population. *J Intern Med* 253, 553-562.

Wood D., DeBacker G., Faergeman O., Graham I, Mancia G. and Pyörälä K. (1998) Prevention of coronary heart disease in clinical practice. Summary of recommendations of the second joint task force of European and other societies on coronary heart prevention. *Eur H J* 80, 1-10.

Summary in Swedish – Sammanfattning på svenska

Läkemedelsbehandling vid förhöjda kolesterolvärden är ofta befogad när risken för framtida hjärt-kärlsjukdom är hög och som komplement till livsstilsförändringar. Gällande behandlingsrekommendationer (guidelines) föreslår en uträkning av den numeriska risken som underlag för att avgöra vilka av tidigare hjärt-kärlfriska individer som skall erbjudas behandling. Guidelines och hjälpmedel för riskuträkning är ofta inte tillgängliga vid patientkontakten och tycks användas i liten omfattning, vilket kan leda till över- och underbehandling med läkemedel.

Syftet var att undersöka allmänläkares beslutsfattande om läkemedelsbehandling för patienter med hyperkolesterolemi. Patientfall presenterades antingen i pappersformat (studie I och II) eller på datorskärm (studie III-V). I studie I skattade 38 allmänläkare sin benägenhet att rekommendera läkemedelsbehandling för 40 fallbeskrivningar. Den enskilde läkarens bedömningsstrategi definierades som uppsättningen av statistiska regressionsvikter för de variabler som användes för att beskriva patientfallen. Strategierna skiljde sig avsevärt mellan läkare. Den viktigaste variabeln var tidigare kranskärlssjukdom och näst viktigast var graden av kolesterolförhöjning. De flesta tycktes använda två eller tre av de nio presenterade variablerna vid bedömningarna. I studie II undersöktes allmänläkares och medicinstuderandes uppskattningar av patienters risk för framtida kranskärlssjukdom. Båda grupperna visade en tendens att underskatta risk, speciellt för högriskpatienter.

I studie III-V användes tänka-högt teknik för att undersöka tankeprocesser bakom beslutet att rekommendera läkemedelsbehandling eller inte. Allmänläkarna instruerades att högt återge alla sina tankar om den information om patientfallen som presenterades på en serie datorskrämar. Allt som sades togs upp band. De utskrivna protokollen kodades sedan för om de separata uttalandena talade för eller emot läkemedelsbehandling (beslutets riktning). De kodades också för vilken information om patienten som åsyftades och för typ av kognitiv aktivitet. I studie III fick hälften av deltagarna inte bara tänka högt utan också skatta benägenheten att förskriva läkemedel efter varje skärm med ny information om patientfallet. Skattningar och tänka-högt protokollen återspeglade förändringen i beslutets riktning på likartat sätt, vilket tolkades som ömsesidigt stöd för validiteten i de två svarsmåtten. Tänka-högt protokollen återspeglade minst lika väl som skattningsdata denna förändring av ett besluts riktning över tid.

I studie IV och V analyserades tänka-högt data med avseende på hur ofta olika slags information om patientfallen värderades och hur de användes vid besluten. Graden av kolesterolförhöjning värderades oftast och olika bedömare verkade använda olika kritiska nivåer för vad som var betydelsefull grad av kolesterolförhöjning. Livsstilsfaktorer tycktes bedömas utifrån olika perspektiv. Att en patient var rökare kunde av vissa allmänläkare betraktas som en riskförhöjande faktor vilket därmed också ökade benägenheten att förskriva läkemedel medan andra läkare tycktes se rökstopp som ett alternativt angreppssätt och således som ett argument att avstå från eller avvakta med läkemedelsbehandling.

Efter de sex patientfallen ombads läkarna att beskriva med egna ord hur de brukar resonera kring sina patienter med kolesterolförhöjning. Dess protokoll kodades för orsak-verkan samband för att få en bild av enskilda läkares kunskap och åsikter kring vad som är viktigt för förskrivningsbeslutet. Allmänläkarna tycktes i flera fall göra avvikelser från en strikt tillämpning av guidelines, trots att de hade kunskap om de relevanta delarna av guidelines av de verbala protokollen att döma. I dessa fall var argumenten ofta livsstilsfaktorer.

Tänka-högt metodiken med det kodningsschema som utvecklades i studie III-V förefaller vara ett värdefullt komplement till statistiska angreppssätt (som användes studie I och II). Ett exempel på tillämpningsområde är att studera skillnader i beslutsprocesser mellan erfarna läkare och medicinstuderande för att användas i undervisning. Ett annat tillämpningsområde är att undersöka användbarheten av olika utformningar av beslutsstöd som integreras i den datoriserade patientjournalen.