ISM skriftserie Nr. 48

THE POTENTIAL FOR REDUCING INAPPROPRIATE HOSPITAL ADMISSIONS:

A study of health benefits and costs in a department of internal medicine

Bjørn Odvar Eriksen

Tromsø 1999

Institute of Community Medicine
University of Tromsø, Norway

Department of Internal Medicine
University Hospital of Tromsø, Norway

ISM skriftserie
blir utgitt av Institutt for samfunnsmedisin
Universitetet i Tromsø.

Forfatterne er selv ansvarlige for sine funn og konklusjoner. Innholdet er derfor ikke uttrykk for ISM's syn.

The opinions expressed in this publication are those of the authors and do not necessarily reflect the official policy of the institutions supporting this research.

ISBN 82 - 90262 - 55 - 8
1999

TABLE OF CONTENTS

Acknowledgements

This thesis is based on the Tromsø Medical Department Health Benefit Study, which was initiated in 1991 by Jan Fr. Pape during his time as chief physician at the University Hospital of Tromsø. A project group consisting of Ivar Sønbø Kristiansen, Erik Nord, Jan Fr. Pape and myself designed and planned the study and was responsible for preparing the five papers presented here. The thesis was composed with Olav Helge Førde and Ivar Sønbø Kristiansen as my supervisors. I am deeply indebted to all these contributors for their continuous support.

Sven M. Almdahl, Anne Hensrud, Steinar Jæger, Fred A. Mürer, Reidar Robertsen and Glen Thorsen participated in the expert panels and spent countless hours assessing hospital stays as well as taking part in preparing the manuscripts. This study would not have been possible without their efforts.

My colleagues at the Section of Nephrology, Joar Julsrud and Markus Rumpsfeld, made it possible for me to be relieved of my clinical duties to work on the project, and the Department of Internal Medicine gave me research leave. The Norwegian Medical Association Funds for Quality Improvement and the University Hospital of Tromsø Research Fund financed the study. I am very grateful for all these contributions.

Last, I want to thank my most enduring supporters: Siri, Ane, Eivind and Margareth.

List of abbreviations

DRG	diagnosis related groups
HSQG	health-related short-term quality of life gain (definition, see Table 2)
HYE	healthy-years equivalents
ΔHYE	gain in healthy-years equivalents
LEG	life expectancy gain (definition, see Table 2)
LQG	long-term quality of life gain (definition, see Table 2)
NHSQG	non-health-related short-term quality of life gain (definition, see Table 2)
NOK	Norwegian kroner
QALY	quality-adjusted life years
R^2	coefficient of determination
RCT	randomised controlled trial
USD	US dollar (all conversions made according to the exchange rate USD 1 = NOK 7.50)

List of papers

Paper 1. Eriksen BO, Almdahl SM, Hensrud A, Jæger S, Kristiansen IS, Mürer FA, Nord E, Pape JF, Robertsen R, Thorsen G. Assessing health benefit from hospitalization: Agreement between expert panels. **Int J Technol Assess Health Care** 1996; **12**: 126-135.

Paper 2. Eriksen BO, Kristiansen IS, Nord E, Pape JF, Almdahl SM, Hensrud A, Jæger S, Mürer FA, Robertsen R, Thorsen G. Does admission to a medical department improve patient life expectancy? **J Clin Epidemiol** 1997; **50**: 987-995.

Paper 3. Eriksen BO, Kristiansen IS, Nord E, Pape JF, Almdahl SM, Hensrud A, Jæger S, Mürer FA, Robertsen R, Thorsen G. Does admission to a department of internal medicine improve patient quality of life? **J Intern Med** 1998; **244**: 397-404.

Paper 4. Eriksen BO, Kristiansen IS, Nord E, Pape JF, Almdahl SM, Hensrud A, Jæger S. The cost of inappropriate admissions: A study of health benefits and resource utilisation in a department of internal medicine. **J Intern Med** 1999; (In press)

Paper 5. Eriksen BO, Førde OH, Kristiansen IS, Nord E, Pape JF, Almdahl SM, Hensrud A, Jæger S, Mürer, FA. Cost savings and health losses from reducing inappropriate admissions to a department of internal medicine. 1998; (Submitted)

Summary

Inappropriate hospital admissions are defined as those which do not result in health benefit for the patient or in such benefit that could have been obtained on a lower care level. Studies from many parts of the world have reported high rates of such admissions. It is commonly believed that they represent a potential for significant cost reductions. However, this assumes that they can be identified at the time of admission, and, furthermore, that their cost is comparable to that of appropriate admissions. These assumptions were investigated in the Department of Internal Medicine at the University Hospital of Tromsø.

At present, any intervention to reduce inappropriate admissions would have to involve clinical judgement in one way or another. To explore the feasibility of this approach, two panels of experienced clinicians used a two-round structured consensus method for assessing the appropriateness of consecutive admissions to the department during a six-week period. Both panels consisted of an internist, a surgeon and a general practitioner, who were all board-certified. The panels first tried to predict the appropriateness of the admissions solely from the information available at the time of admission. After discharge, they then made a final judgement of appropriateness with the additional information collected during the stay. To avoid bias, one panel made the predictions and the other panel the final assessments for half of the admissions, and vice versa for the other half. The assessments of appropriateness were based on estimates of gains in life expectancy and quality of life, and of necessary care level. The direct costs to the hospital of each stay were estimated.

To explore the agreement between the panels, a 10% random sample of the included admissions was assessed by both panels after discharge. This demonstrated reasonable agreement about the assessments of health-related benefits and appropriateness.

As judged by the panels after discharge, about one quarter of the admissions were inappropriate. The health benefits were very unevenly distributed across the appropriately admitted patients. A few patients had gains corresponding to life-

during or shortly after the hospital stay. The mean cost of the inappropriate admissions was less than half that of the appropriate, and the inappropriate admissions only accounted for 12% of the total costs.

When trying to detect inappropriate admissions on the basis of the information available at the time of admission, the panels performed poorly. Only about a quarter of the admissions judged inappropriate in the final assessment after discharge was identified. About a tenth of the appropriate admissions was falsely classified as inappropriate. The savings from denying care for admissions considered inappropriate at admission, would have been modest. Health losses would have occurred because some patients with health benefits would not have been admitted. Compared to other medical interventions accepted as cost-effective, the potential cost savings were small relative to the potential health losses.

It is concluded that efforts to reduce inappropriate admissions based on predictions of health benefit and care level at admission are unlikely to result in savings that will justify the health losses.

1. Introduction

The purpose of health care is to attain a state of "complete physical, mental and social well-being and not merely the absence of disease or infirmity", as ambitiously formulated in the World Health Organisation's definition of health [1]. Although it is recognised that health care is less important for this goal than political and social conditions, the expenditures for health care are increasing steadily in all parts of the world [2]. In Norway, they rose from 4.6 % of the gross domestic product in 1970 to 7.9 % in 1996. The corresponding percentages for the USA were 7.2 and 14.2, which were the highest in the world.

Parallel to this development, progress in medical science is creating a higher demand for health care. A gap is emerging between what is technologically possible and what society can afford. Already, in many situations, limited resources rather than limitations in medical technology decide what can be done for patients. This challenge has been met with various strategies for efficiency improvement, priority-setting and cost-containment. While there is debate about which strategies should be implemented, all seem to agree that health care that does not improve health should be reduced as much as possible, and that health care resulting in only marginal health benefits may have to be rationed. It is widely believed that the reduction of so called ineffective or inappropriate health care could lead to considerable savings which would allow us to provide better care for other patients [3-5].

This belief provided the starting point for the Tromsø Medical Department Health Benefit Study, which focused on inappropriate hospital admissions to a department of internal medicine. The study was designed to investigate the relationship between health benefits and costs, and to estimate the potential for savings by reducing the number of admissions.

2. Background

2.1 The effectiveness of medical care

Research on inappropriate health care must be seen within the context of the debate about the effectiveness of medical care in general, which has been going on since the 1950s. Some of the most extreme critics of medical care have claimed that it has very little or even a negative effect on population health [6-10]. These claims were supported by studies which indicate that health care has less effect on mortality and morbidity than political and social factors [11-15], and by more anecdotal evidence like the reduction in mortality observed during a doctors' strike [16]. Others have vigorously defended medical care against these attacks [17-22]. Studies have shown that the mortality from diseases amenable to medical treatment has declined [23-25], and the point that mortality may be a poor indicator of the effects of modern medicine has been made [26-28].

An important implication of the criticism of medical care is that much of it may be unnecessary or inappropriate. Studies of different kinds of inappropriate health care will be reviewed in the following sections.

2.2 Inappropriate health care - terminology

There is no universally accepted definition of the term "inappropriate health care" in the literature although it is commonly used about care that provides no health benefit for the patient. Other terms are used for specific instances of such care, as e.g. "inappropriate hospital days". Though the subject of this thesis is inappropriate admissions to hospitals, literature on both inappropriate admissions and other related kinds of inappropriate health care will be reviewed.

2.3 Geographical variations in health care

Many studies have shown that there are geographical variations in care which cannot be readily explained by medical factors. These variations have been taken as indicators of inappropriate health care. The most important have been studies of variation within small geographic areas. In a series of studies from USA, Wennberg et al found large variations in hospital days, hospital discharges and surgical procedures per 1000 persons in Vermont [29], and, later, in health care expenditures

per capita between Boston and New Haven [30]. These differences were not associated with different mortality rates in the two areas [31]. Other investigators have found significant variations in the rates of surgical procedures [32,33] and in the rates of both medical and surgical procedures [34]. The same findings have been demonstrated in other developed countries [35]. Geographical differences in the utilisation of hospital care, rates of surgical procedures and use of perioperative total parenteral nutrition have also been reported from Norway [36-38].

It is difficult to explain these variations by differences in the incidence of disease [39]. Rather, it has been hypothesised that they are caused by differences in physicians' practice patterns. The greatest variations have been found for conditions for which there is disagreement about diagnosis and treatment [29,30,34]. This has been called the "professional uncertainty hypothesis" [40], and more research is called for to fill the gaps in medical knowledge that presumably are the most important causes of the variation [41].

A recent study investigated small area variation for a longer time period and compared different methods of analysing the data. It was shown that the magnitude of the variations depended on both the method of analysis and the time period, indicating that the small area-variation in hospitalisation rates may vary substantially less than has been previously reported [42].

2.4 Inappropriate medical interventions

It has been discussed to what extent the small area variation in care is caused by inappropriate use of interventions for which there is reasonable agreement about the indications. This issue has been explored by the RAND-UCLA Health Services Utilisation Study (HSUS), where consensus about appropriate indications for six medical and surgical procedures was developed by panels of expert physicians [43-45]. Using these criteria, the authors found that 17% of the coronary angiographies, 32% of the carotid endarterectomies and 17% of the upper gastrointestinal tract endoscopies were inappropriate, but that this could not explain the geographic variation in the use of these procedures [46]. Another study by Leape et al reached the same conclusion [47].

The rates of inappropriate procedures reported in other studies vary. In the UK, the rates for cardiovascular procedures were similar to those in the RAND-UCLA HSUS

[48]. A study of second opinion of coronary angiographies for patients with stable angina pectoris in Boston indicated that 50% of these procedures were unnecessary or could be postponed [49]. In New York, lower percentages of inappropriateness were found in a later study of coronary angiography (4%) [50], coronary artery bypass graft surgery (2.4%) [51] and percutaneous transluminal coronary angioplastys (4%) [52], but many procedures were carried out on uncertain indications (20%, 7% and 38% respectively). Similar low rates of inappropriateness have been reported from a consortium of academic medical centres in the USA [53] and in Sweden [54].

Though some see these studies of inappropriate health care as heralding a revolution in health care [4], others have attacked their methodological foundations [55]. The theoretical basis of the "appropriateness" concept has been questioned [56] as well as the method used for establishing appropriate indications [57]. Data on the method's sensitivity and specificity have been called for, and it has been suspected that the rate of procedures falsely judged inappropriate may be high [58].

2.5 Inappropriate hospital days

Other investigators have focused on inappropriate days spent in hospital. This was the purpose of the Appropriateness Evaluation Protocol, which was developed in 1981 by Gertman and Restuccia [59]. The implication was that such days could be eliminated and costs reduced. An inappropriate hospital day was defined as one where..

"...patients receive either services that provide no significant benefit or services that could be rendered in a less costly lower-level institutional or outpatient setting."[59]

The Appropriateness Evaluation Protocol is a screening tool consisting of a diagnosis-independent set of criteria for review of medical records by nurses. It has been validated both against the judgement of one physician reviewer [59] and panels of physicians [60,61], and has been found reliable in several studies [59-66]. Using this instrument, rates of inappropriate hospital days ranging from 8 to 15% in Switzerland [67], 28 to 49% in Italy [68], 15 to 44% in Spain [69], 46% in Portugal [70], 12 to 39% in USA [60,71] and 29% in South Africa [63] have been found in different types of hospitals and departments.

2.6 Inappropriate hospital admissions

The first studies of hospital admissions characterised as "inappropriate" or judged to confer the patient no health benefit appeared towards the end of the 1950s. One of the earliest was Crombie and Cross' study of patients in the medical wards of a Birmingham general hospital. They concluded that a quarter of them had "...no diagnostic or therapeutic requirements at hospital level" [72]. Several studies were carried out in the 1960s, finding percentages of inappropriate admissions ranging from 1.6 to 50 in different types of hospital departments [73,74]. Most of these studies used some form of physician judgement for detecting inappropriateness.

In the USA, concern over the rapid rise in expenditures for the Medicaid and Medicare programs led to the establishment of Professional Standards Review Organisations for performing utilisation reviews in hospitals by the early 1970s [75]. These reviews included studies of inappropriate admissions. After several attempts to find a reliable method for detecting such admissions, the Appropriateness Evaluation Protocol was developed by Gertman and Restuccia in 1981 [59]. Originally, the protocol was only intended to measure inappropriate days of hospital stays, as mentioned above, but was later established as a tool for admissions as well [66](Table 1). It has been validated against the judgement of physician reviewers [60,61]. At about the same time, other similar sets of criteria were developed [61], but have been less used than the Appropriateness Evaluation Protocol, judged from the number of published studies.

In 1987, Payne reviewed the results of investigations of inappropriate admissions in the USA [71]. She found percentages of inappropriate surgical and medical admissions ranging from 10 to 40 in studies that had used the Appropriateness Evaluation Protocol. In the late 1980s, the Managed Care Appropriateness Protocol was developed on a commercial basis by one of the originators of the Appropriateness Evaluation Protocol [75]. In a MEDLINE search, I was not able to find any studies using this instrument.

Since 1990, few studies of inappropriate admissions in the USA have been published. On the other hand, there are numerous studies using the Appropriateness Evaluation Protocol from other parts of the world. Six per cent inappropriate admissions were found in a teaching hospital in Australia [76]. In Europe, the protocol has been used in a co-ordinated effort to assess inappropriate admissions

by the European Union [77,78]. Using various modifications of it in different kinds of hospitals and departments, percentages of inappropriate admissions ranging from 1.3 to 25 were found in Portugal, Spain, Italy, Switzerland and France [67-70,79]. In Britain, the Bed Study Instrument, a closely related instrument, has been used in some studies where low rates of inappropriate admissions have been found, in one study less than 1% [80]. This study has later been criticised on methodological grounds [81]. A recent study from the UK found approximately 20% inappropriate emergency admissions to a department of general medicine and care of the elderly [82].

In Norway, three studies have assessed the appropriateness of admissions to departments of internal medicine as judged by the attending physicians. Mosvold et al investigated whether admissions to Aker Hospital in Oslo could have been avoided. This was found to have been the case for 19% of the admitted patients [83]. In a later study from the same hospital, the percentage had dropped to 4 [84]. Even though the first study was referred to in a publication from the latter [85], the difference between the two studies was not commented. Øie et al let discharging physicians assess whether admissions to the department of internal medicine at the Diakonissehjemmet Hospital in Bergen could have been avoided. This was found to have been the case for 42% [86]. An observation unit was specified as one of the alternatives to admission. If stay in such a unit had been defined as an admission, the percentage would have decreased to 20.

These three studies used a form of clinical judgement called implicit review, i.e. explicit criteria were not used [71]. Three other studies from the northern part of Norway did use criteria for health benefit to evaluate admissions to departments of internal medicine at local hospitals. Sander found that 48% of the patients admitted to Kirkenes Hospital did not achieve any benefit as judged by one reviewer according to three criteria for benefit [87]. From Narvik Hospital, Seip et al reported that only 1 of 600 patients did not benefit [88]. This study did not specify its method for assessment of benefit. In 1983, Syse et al repeated Sander's investigation at Kirkenes Hospital. They included a study of the agreement between two reviewers using Sander's criteria and found a kappa-statistic of 0.53. Thirty-five per cent of the patients experienced no benefit [89], which was not very different from Sander's result.

2.6.1 Factors associated with inappropriate hospital admissions

In 1987, Payne reviewed factors associated with inappropriate hospital admissions [71]. She found that none of several patient characteristics tested in the reviewed studies were consistently associated with inappropriate admissions. Siu et al investigated the effect of cost sharing by the patient on the rate of inappropriate admissions. No significant association was found, but cost sharing was found to reduce the rate of appropriate hospitalisation [60]. In the same study, there was a significant higher percentage of inappropriate admission of women [90]. In a more recent study, Perneger et al studied factors in the patient's social situation associated with inappropriate admissions in Switzerland. Better physical functioning of the patient, lower mental health status of the patient's spouse, receipt of informal help from family or friends, and hospitalisation by one's own physician, were found to predispose [91].

No hospital characteristic has consistently been found to be associated with inappropriate admissions, though one study did find an association between inappropriate admissions and number of beds in the hospital, and another with shorter length of stay [71]. Physician characteristics may be important as one study has reported great variation in the rates of inappropriate utilisation among physicians within hospitals, and a significant effect of informational feedback [66,71]. One study also found a higher percentage of inappropriate admissions by physicians licensed for more than 15 years [90].

Three studies of inappropriate admissions according to diagnostic category found wide variation in rates. Diseases of the blood and blood forming organs, myeloproliferative disorders and digestive disorders had high rates in all of the studies, while disorders of the eye, infectious and parasitic disease, pregnancy and trauma had low rates [71].

Two studies did not find any association between admission rates and rates of inappropriate admissions in different geographic regions in the USA [60,92].

2.7 The cost of inappropriate health care

Few studies of inappropriate health care have investigated its cost, but some studies have explored the association between indicators of inappropriate health care and cost. Most of these have been carried out in the USA.

2.7.1 Poor prognosis and high cost

Several studies have examined the relationship between cost and survival from a hospital perspective. Schroeder et al made a follow-up study of the 13% of patients with the highest charges from nine acute-care hospitals in the San Francisco Bay area in 1976. After two years, 34% of these patients had died [93,94]. Zook and Moore examined outcome at discharge in relation to charges for a random sample of patients at six different hospitals in the same year. They found that 20% of the patients with charges greater than the 80th centile died in hospital, as contrasted by 4% of the rest [95]. Pompei et al also found a high mortality for patients with high costs in a medical service at a New York Hospital in 1984 [96].

These studies suggest that the average cost of patients with a poor short- or long-term prognosis is higher than for other patients, which could indicate a waste of resources. This issue has been investigated more directly in several studies from intensive care units in the USA and in some studies from Europe. The reason for the special interest in these departments in the USA is that there are more intensive care unit beds per inhabitant here than in any other country [97].

Detsky et al investigated hospital charges for survivors and non-survivors in a general combined intensive and coronary care unit from 1977 to 1979. He found that the charges of the 9% who died were about the double of those who survived. These 9% incurred 17% of the total charges in the study period [98]. Higher costs for non-survivors in a medical intensive care unit were also found by Fedullo et al for some age groups, but only small differences when the total hospital charges were compared [99]. In 1984, Oye et al found that the high-cost 8% of patients in a medical intensive care unit used as many resources as did the low-cost 92%. The in-hospital mortality was 71% in the first group and 20% in the second, and this difference was statistically significant [100]. Similar results have been reported from studies of intensive care units in Sweden, Germany and Spain [101-103].

High costs have also been found for cancer patients. Schapira found that about three fourths of cancer patients admitted to an intensive care unit spent less than three months at home before dying, and that the cost of one additional life year for this group was USD 82,000 for patients with solid tumours and USD 190,000 for haematological cancers [104]. In Finland, Holli and Hakama found that patients with

breast cancer who died received more treatment than survivors. This study did not include a cost analysis [105].

2.7.2 Medical costs at the end of life

Since the introduction of Medicare and Medicaid in USA in 1965, several studies have explored the relationship between survival and reimbursement by these services. They have uniformly reported that a disproportionately high percentage of the expenditures are used by enrolees in their last year of life [106]. Lubitz and Prihoda found that, in 1978, the 5.9% who died accounted for 28% of Medicare expenditures. Furthermore, the intensity of resource utilisation increased as death approached, so that 46% of the costs in the last year of life were spent during the last 60 days [107]. This pattern has persisted [108,109], and similar findings have been made in other developed countries [110,111]. The implication is that a patient with short survival in spite of high costs has only had small benefit from whatever medical care he has received, and that these resources may have been wasted.

It was generally found that most of the costs incurred in hospitals. However, when the distribution of costs were studied, few of the decedents had had costs that would indicate treatment in intensive care units or similar costly life-supporting treatment [107,109]. Consequently, the high average cost resulted from standard hospital treatment with higher cost than for those who survive. Other studies have shown wide variation with different causes of death, cancer being the most costly [112].

These studies did not investigate whether the care delivered to dying patients was appropriate, i.e. whether these patients experienced improvement in quality of life or at least some gain in life expectancy. In one small study, Scitovsky related the cost of health care in the last year of life to the patients' functional status. She found that though the total cost was not influenced by the patient's degree of impairment, the cost of hospital care was markedly lower for patients with low scores for activities of daily living, instrumental activities of daily living and cognitive status [113]. Instead, these patients incurred higher costs for home health care and in nursing homes. If it is assumed that these impaired patients had a lower probability than others of achieving health benefits from more intensive treatment, these findings indicate that health care for the dying may be more appropriate than is commonly believed.

2.8 Interventions to reduce inappropriate health care

In the USA, different strategies for cost containment in health care have been implemented. Many of these are parts of so-called utilisation review, which "encompasses a broad, heterogeneous group of interventions, most commonly involving the prospective review of decisions to admit patients to hospital and perform certain procedures, but also including concurrent evaluation of inpatient care and of the management of high-cost cases."[114]. Utilisation review has grown into a industry of its own, and its effect on medical care has been explored in several studies. There is some evidence that utilisation review can reduce hospital costs, mostly through reducing the number of admissions [115-117] and by shifting some of the costs to outpatient care [118]. However, as several investigators point out, little is known about how this affects the quality and outcome of care. Even if utilisation review programs are meant to reduce only inappropriate care, one study has suggested that the reduction is more an unspecific effect of being reviewed than of the application of the criteria in the program [119]. This could indicate that appropriate hospital utilisation is being reduced as well.

In at least two studies, the Appropriateness Evaluation Protocol has been part of the utilisation review. One of these found a reduction in the mean percentage of inappropriate admissions to six hospitals from 6.9 to 3.3 after the program had been implemented [74]. However, in a controlled trial, Payne et al found no effect of feedback about the rate of inappropriate admissions to 11 hospitals in Massachusetts, although there was a general decline in the percentage of inappropriate admissions during the study period [120].

In Israel, two controlled studies of interventions to reduce inappropriate hospital days have been performed for medical [121] and paediatric patients [122]. The intervention consisted of the requirement that patients who did not fulfil the criteria of the Appropriateness Evaluation Protocol for an appropriate hospital day should be discharged unless a consultant gave his consent to a continued stay. The studies found a reduction of inappropriate days in both the study and control groups, but greater in the study group. Because of methodological limitations of the studies, these results are difficult to interpret.

Other studies have assessed the potential for savings by reducing unnecessary and excessive care at the end of life. One approach has been to promote the use of

advance directives which are given by patients to avoid futile life supporting treatment against their wishes. However, Teno et al found that an intervention which increased the documentation of such directives did not decrease hospital resource use [123]. In the same study, she demonstrated that the savings from reducing interventions for patients at high risk of imminent death would only be modest [124]. An intervention to save resources through improved communication about preferences for treatment between physicians and patients at high risk of dying was ineffective [125]. Emanuel and Emanuel reviewed the results of these and other strategies for cost reductions at the end of life and concluded that the savings were unlikely to be substantial [126].

2.9 Summary of the literature review

- There is great variation in the rates of inappropriate health care across different studies. This applies both to interventions and to hospital stays. Some of this variation is probably caused by differences in definitions and methods. However, few investigators have reported rates below 10%. High rates have been reported for both surgery and internal medicine and from countries in all parts of the developed world. This indicates that health care which does not result in health benefit for the patients is a serious problem.
- Several studies have shown that a large share of health care resources are used on patients with short life expectancy. This applies both to patients hospitalised in intensive care units and to patients at the end of life. While most of these studies have been carried out in the USA, similar results have been found in several European countries. Since these patients could only have achieved limited health benefit from whatever interventions they had undergone, the implication is that resources were wasted.
- In the USA, rates of hospital admissions and costs have been reduced through various forms of utilisation review, but it has not been convincingly demonstrated that interventions specifically targeting inappropriate admissions can obtain savings. It is also doubtful whether it is possible to save resources for care at the end of life.

2.10 Limitations of previous studies of inappropriate hospital admissions

- Cost analyses to quantify the potential cost reductions represented by inappropriate admissions have not been performed. Some studies have demonstrated substantial rates of inappropriate admissions without investigating cost, while others have studied the cost of treating patients with short survival without making explicit assessments of the appropriateness of the care delivered. Short survival after a hospital stay does not necessarily indicate that the admission was inappropriate, since the patient's life expectancy and quality of life might have been worse without admission.
- To obtain savings by reducing inappropriate admissions, it is necessary that they can be identified as such *before or at the time of admission*, i.e. before the results of diagnostic and therapeutic interventions during the hospital stays are known. None of the referred studies of inappropriate admissions made clear when in the course of the hospital stay the judgement of inappropriateness was made. Several of the criteria of the Appropriateness Evaluation Protocol are based on information that can only be obtained some time after admission (Table 1). It has not been shown that inappropriate admissions can be identified from information available at admission with sufficient accuracy to obtain savings while avoiding health losses from denying appropriate admissions.
- Few studies provide a definition of the term "inappropriate hospital admission". Those who do, base their definition on the concept "health benefit" [59]. I am not aware of any study of inappropriate admissions which has included a description of methods for measuring health benefits. Furthermore, few of the studies use physicians for evaluating admissions, but instead rely on research nurses and screening instruments.

3. Definitions and aims

In this thesis, the following definitions will be used:

- *Health benefit:* Improvement in life expectancy or health-related quality of life from a hospital stay relative to a situation without admission.
- *Inappropriate admission:* A hospital admission which does not result in health benefit, or which results in health benefit that could have been obtained on a lower care level.

The aims of the study were:

- Primary aims

To investigate to what extent clinical judgement based on information available at the time of admission can be used for identifying inappropriate admissions to a department of internal medicine

To explore whether clinical judgement can in principle be used for reducing inappropriate admissions and department costs without, at the same time, resulting in unacceptable health losses.

- Secondary aims

To estimate the proportion of inappropriate admissions, and the share of the total costs that they represent, in a department of internal medicine.

To study the agreement between expert panels for assessments of health benefit.

4. Methods

4.1 Study population

During a six week period from 1st February 1993, all 521 admissions to the department of internal medicine at the University Hospital of Tromsø were eligible for inclusion in the Tromsø Medical Department Health Benefit Study (Figure 1). Patients are sometimes transferred for administrative reasons after having been treated in other university hospitals. These patients were excluded (n=3), as were also patients admitted to the clinical research unit (n=2) and one patient whose medical record could not be found. Readmissions occurring in the study period that had been scheduled during a stay prior to the study period were excluded (n=27). Most of these patients were admitted for evaluation or continuation of treatment. Readmissions in the period that had been scheduled during a previously included stay were merged with the primary admission (n=9). The number of admitted patients was 462 of whom 17 had 2 separate included admissions, i.e. they had one un-scheduled readmission in the study period. Accordingly, 479 admissions were included in the study.

4.2 Design

Two expert panels were recruited, each consisting of an internist, a surgeon and a general practitioner. Using a consensus method with two rounds, they estimated the health benefit and appropriateness from each admission after discharge on the basis of comprehensive summaries of all relevant information about the patient. To investigate the panels' ability to predict the health benefits, exactly the same assessments were also made only from information available at admission. In the following, the former will be termed "discharge assessments" and the latter "admission assessments". Bias from letting the same panel make both assessments for the same admission was avoided by using two panels. Each panel assessed half of the patients at admission and the other half after discharge and vice versa (Figure 2). The patients were randomised to each half using a random number generator. A 10% random sample was drawn to study the agreement between the panels. These patients were assessed by both panels after discharge.

4.3 Assessment of health benefit and necessary care level

4.3.1 Health benefit

Benefits from hospital stays can be classified according to different criteria. One distinction can be made between benefits for the patients and benefits obtained by other persons or society, e.g. relatives. One example is the benefit to society from isolating a patient with a communicable disease. For practical reasons, we only assessed benefits experienced by the individual patients.

Another distinction can be drawn between health-related benefits and benefits which are unspecific effects of hospitalisation, as e.g. the provision of shelter for a homeless person or relief from a difficult social situation. The latter type of benefit can be defined as not resulting from specific medical treatment or care. In the present study, an attempt was made to study both kinds of benefit.

A third distinction is usually made between life expectancy gain (LEG) and gain in quality of life. In this study, both were estimated and the results presented in Papers 2 and 3. The time trade-off method was used for estimating gains in quality of life [127]. This technique gives a measure of quality of life ranging from 0 (corresponding to death or coma) to 1 (corresponding to full health). Its validity and reliability have been established by others [128]. The questionnaires used for the assessments are included in Appendix 1 and 2, and the details of the procedure explained in Paper 1.

In the literature, different reference groups have been used for measuring the quality of life of health states. It has been discussed which groups should be used, since it is known that the assessments of patients and e.g. physicians may differ [129]. Torrance says that "..The answer can be determined, in part, from the purpose and the viewpoint of the study...." [127]. We wanted to examine the relationship between use of resources and physicians' assessments of health benefit, which justifies the use of physicians' assessments of quality of life in this study. A more detailed discussion of this issue can be found in Paper 1.

The two dimensions of health benefit can be combined in a measure of life expectancy adjusted for quality of life. The most common of these measures is the *quality-adjusted life years* (QALY) [130], which measures health outcomes as a product of gain in quality of life and the number of life years that the patient get to enjoy the health benefits. One QALY is equivalent to one year in perfect health.

Although the use of QALY for prioritising has been criticised [131-136], this measure is now widely used. In the present study, the closely related measure *healthy-years equivalents* (HYE) was used [137]. The properties of HYE relative to QALY are subject to controversy [138-142]. The main difference is that while QALY are based on health benefits in individual years being valued one by one and then added together, HYE derive from holistic valuations of life scenarios. The latter approach allows the valuator to take into account dependence between life years (contextual effects), for instance the effect of prognosis. However, because it is more well-known, the term "QALY" was used in Papers 4 and 5.

In principle, all benefits from hospital stays could have been measured in HYE. However, our application of the time trade-off method had limited sensitivity for benefits of short duration relative to the remaining lifetime of the patient. To illustrate this, consider patients treated for chronic renal failure with hospital hemodialysis, who have been found to have a quality of life around 0.50 in different studies (range 0.41 to 0.58) [143-147]. A gain of 1 HYE would be equivalent to restoring the life of one such patient to full health for 2 years, which is a considerable gain. At the other end of the scale, consider a patient with pain from a gallstone attack which has limited duration and can be effectively treated with drugs. This patient will probably have a very low quality of life during the attack, e.g. 0.2, but because of its short duration (e.g. 5 hours), the gain from treating the pain with analgesics could not exceed 0.0005 HYE, which would have been the gain if the quality of life had been raised from 0.2 to 1.0 while the attack lasted.

It would have been very difficult to measure this gain with our application of the time trade-off technique. For this reason, it was necessary to measure quality of life gains during or shortly after the hospital stay on an ordinal scale. Separate scales were used for health-related short-term quality of life gain (HSQG) and non-health-related short-term quality of life gain (NHSQG). The definitions of the different types of health benefit estimated in this study can be found in Table 2 and more details in Paper 3.

Diagnostic interventions sometimes result in a diagnosis without any improvement in the patient's health. However, many patients will experience relief by being provided with an explanation of symptoms and other manifestations of disease. The elimination of a tentative diagnosis proposed by the referring physician may have the

same effect. At least one study has shown that patients may be willing to pay for diagnostic information, even if it does not result in any specific medical consequences [148]. In the instructions to the expert panels, this type of improvement in quality of life was explicitly defined as NHSQG.

It may also be argued that obtaining new diagnostic information should be regarded as a benefit, regardless of whether it improves the quality of life of the patient. It is certainly of value to physicians and the hospital to be able to solve diagnostic problems, especially from the perspectives of research and education. The clinical competence gained will benefit future patients. Since this study only aimed to assess benefits for the included patients, we did not count diagnostic gains alone as a benefit. However, the experts did assess whether a diagnosis that could explain disease manifestations had been made.

4.3.2 Health benefit attributable to the hospital stay

To find the gain in HYE (ΔHYE) attributable to the hospital admission, it is necessary to consider the patient's situation in the event that he had not been admitted or treated elsewhere for his current health problem. The experts therefore made a separate assessment of the patient's expected remaining HYE for this situation, and the gain attributable to the hospital admission was found by subtracting this amount from the HYE expected after the stay. Assuming that the patients would not have received treatment if not admitted is unrealistic, since many of them would then have been treated on a lower care level. This was taken into account by making a separate assessment of the care level necessary to obtain the health benefit (see next section).

For patients with chronic conditions, one hospital stay may be only one of several care episodes which occur over many years and which include treatment in other departments as well as consultations in the outpatient clinic. All these separate episodes can be said to be *necessary*, but none *sufficient* alone, for obtaining the health benefit that ultimately results from the patient's contact with the hospital. E.g. a patient with diabetes mellitus will have appointments for routine follow-up in the outpatient clinic as well as stays for complications as diabetes coma, diabetic nephropathy, diabetic retinopathy etc. The patient will probably achieve health

benefit from his contact with the hospital, but there is no obvious way to apportion this benefit to each of the care episodes.

In the present study, we defined the health benefit from an included stay as the health benefit for which it was a *necessary* condition. This means that if the stay was a part of a prolonged contact with the hospital for a chronic condition, the assessment of health benefit took into account the expected effect of future care for the same condition. E.g. if a patient with end-stage renal disease was admitted for the initiation of hemodialysis, the assessment of benefit from this stay presupposed that the treatment would continue after discharge.

4.3.3 Care level

The concept of appropriate level of care plays a central role in the cost containment debate. The experts were asked to consider whether patients with health benefits could have obtained the same gain on a lower care level. They were given the choice between "primary care", "outpatient clinic" and "hospital". This was done as a part of both the admission and the discharge assessments.

The University Hospital of Tromsø is a tertiary referral hospital, and many of the elective patients are referred from local hospitals which represent a lower care level. In this study, no attempt was made to distinguish between different levels of hospital care.

4.3.4 Data

Both the admission and discharge assessments were made from summaries of clinical information which were based on the medical record at admission and the discharge reports respectively. To ensure that the admission summaries should be as complete as possible, the project co-ordinator (B.O.E.), who is a board-certified specialist of internal medicine, checked the medical records for completeness and obtained missing information from the physicians and nurses in charge of the patient. No information was deleted in the editing process, so that the full text of the medical record was contained in the summary. The summaries were prepared before the admissions were randomised. They were blinded for data which could identify patient or physician.

After discharge, the same procedure was applied to the discharge reports. Information about planned interventions within two months after discharge was included. It could be argued that a longer period would have improved the estimates. While this is certainly true for patients with acute or subacute conditions, only observation till death would have sufficed for some of the patients with chronic diseases. Since the length of any period would have been arbitrary, we chose to make the estimates as soon as possible after discharge.

Both the admission and the discharge summaries typically consisted of from 1 to 2 typewritten A4 pages. To avoid confusion, the admission summaries were printed on yellow and the discharge summaries on green paper.

4.3.5 Expert panels and the consensus method

As explained above, the two expert panels made the same assessments at admission and after discharge, with the difference that information collected during the stays was available only for the discharge assessments. To investigate whether the two panels could reach a reasonable level of agreement for making these judgements, a random sample was drawn for assessments by both panels after discharge (Figure 2). The panels were blinded to which admissions were included in this sample. Results of the agreement study were reported in Papers 1, 3 and 5.

The consensus method has been used for a wide range of problems where it is difficult to obtain data by other methods. Its use in medicine has been reviewed by others [149-151]. Our application of the method is a modification of the nominal group technique. Hotvedt et al used a similar method in their study of the benefit of helicopter evacuation, which also included estimates of gains in life expectancy and quality of life [152].

Instead of using three internists, two specialists from other disciplines were chosen. The experts were required to fulfil three criteria: They should

- be board-certified specialists with long practical experience
- not have any affiliation with the investigated department
- practice in northern Norway, so that they would be acquainted with the conditions of health care in the region served by the department.

The experts cannot be considered to be representative of their respective disciplines in the sense that they were not randomly drawn from the population of all possible experts. Rather, they were chosen because it was believed that they would be especially capable of making the estimates required by the study. The justification for this was that the study aimed to investigate whether the prediction of inappropriate admissions was possible in principle.

4.3.5.1 Instructions to the experts

At the start of the study, the experts were convened for a thorough review of the study protocol. The assessment method, and in particular the time trade-off method, was explained in detail and discussed. The experts also received written instructions (see Appendix 3).

In some studies, experts have been given reviews of relevant literature, as e.g. in the RAND-UCLA Health Services Utilisation Study for establishing appropriate indications for different procedures [43]. This was not possible in our setting, since it would have involved literature from the entire field of internal medicine. Instead, the experts were instructed to use the best evidence available in each case: randomised controlled trials when possible, other empirical evidence or, as a final resort, pure clinical judgement.

For the assessments of quality of life with the time trade-off method, it was emphasised that the experts should use the instrument as if they themselves had been in the patient's situation, as opposed to making assumptions about what the response of a particular patient might have been.

4.3.5.2 Pilot study

A pilot study of 10 admissions was performed. This study, which only consisted of one round, confirmed that the experts mastered the assessment method. From its result, minor changes to the instruction manual were made (Appendix 3).

4.3.5.3 First round

Essentially the same procedure was used for the admission and discharge assessments. The admission summaries were sent to the experts within 24 hours

after admission, and the discharge summaries when the results of all diagnostic interventions were available.

In the first round, the experts made their individual assessments at home. Agreement in the panel about a particular admission was defined to exist when all the following predefined criteria were satisfied:

- the difference between the maximum and minimum LEG estimates did not exceed 25% of the average estimated life expectancy of the patient after the hospital stay,
- the difference between the maximum and minimum LQG estimates did not exceed 0.20
- the HSQG, NHSQG and care level assessments did not differ by more than one category
- the assessments of diagnostic gain were identical

The cases with disagreement were selected for the second round. For both panels combined, this amounted to 90% of the admission and 84% of the discharge assessments.

4.3.5.4 Second round

In the second round, the three experts of each panel met to discuss the cases with disagreement. These meetings were led by the project co-ordinator (B.O.E.), who did not take part in the discussions except to clarify issues related to the protocol. The admission and discharge assessments were discussed in separate sessions.

At the beginning of the discussion of each case, the project co-ordinator stated which type of disagreement existed and the assessment of each expert. The experts then read the summary, and the expert with the most deviating estimate gave the reason for his or her assessment. The case was then discussed. An attempt was made first to reach agreement about the patient's prognosis in medical terms, and then about the estimates of health benefit. At the end of the discussion, the experts were given the opportunity to revise their estimates. The median was taken to represent the panel's assessment whether agreement was reached or not. For both panels combined, there was still disagreement according to one or more criterion for 42% of

the admission and 32% of the discharge assessments at the end of the second round.

In each meeting, the cases were discussed in random order to avoid bias from temporal changes in the experts' estimates as far as possible.

Because agreement had to be reached for several measures, a high percentage of the cases had to be discussed in round 2. During these discussions, the experts met in Tromsø for 2 to 3 days at a time. Discussions began at 8 AM and often continued till 4 PM.

4.4 Cost analysis

Direct costs from the hospital's viewpoint during the included stays were estimated using the principles outlined by Drummond et al [153]. When two or more stays were merged, the costs of all the stays were included. Data for the analysis were obtained from the hospital's annual report [154], from the hospital's computerised account system, from various clinical databases and from the medical records.

4.4.1 Capital and depreciation costs

Capital costs are not routinely included in the hospital accounts and were not incorporated in this analysis. Neither does the hospital calculate depreciation costs of equipment, but lists its cost when it is purchased. These costs were included, but because considerable variation from year to year can be expected, they were averaged over the years 1992, 1993 and 1994 for each department . We are aware that this approach may over-estimate depreciation costs because it also includes investment in new equipment in addition to renewal of the old. However, the method was chosen because data for calculating the true depreciation costs were not available. Since the costs of equipment, renewal and maintenance were only 3.1% of the hospital's total costs and 0.7% of the costs of the department of internal medicine, the error made from using this method was small.

4.4.2 Research and education

Research and education are integral parts of the activities of a university hospital and may contribute to higher costs than in other hospitals. It was not possible to estimate these costs separately at the level of the cost centres, and consequently, these costs

were included in the calculation of unit costs for the patient-related services. Since research and education of personnel are necessary to produce these services, this seems justified.

4.4.3 Cost centres

Each clinical service department and each clinical department except the department of internal medicine were defined as separate cost centres. Each ward of the department of internal medicine, including the outpatient clinics, geriatric day care centre and coronary care unit were considered as separate cost centres, as was also the intensive care unit. The intensive care unit is a part of the department of anaesthesiology.

4.4.4 Step down allocation of overhead

The step down allocation method with iterations was used to allocate overhead costs to the cost centres, both for the hospital as a whole and for the allocation of overhead costs within the department of internal medicine [153]. The allocation basis was number of employees, square footage, number of admitted patients or number of patient-days as appropriate.

In the accounts, physician salaries were included in the overhead costs of the department of internal medicine. These costs were deducted from the overhead costs and allocated to the cost centres in the department according to the actual assignments of physician labour in 1993. The same approach was used for physician labour in the intensive care unit.

4.4.5 Estimation of unit costs

4.4.5.1 Clinical service departments

4.4.5.1.1 Radiology, clinical chemistry, microbiology, immunology/haematology, pharmacology, pathology, gastroenterology, dialysis

In the Norwegian health care system, inpatients are not charged directly, and the calculation of the costs of services by these departments was based on the fee schedule for outpatients. The total charges in 1993 of each department according to this schedule were calculated as if all patients had been outpatients, and the cost-to-charge ratio was found by dividing the department's total costs after allocation of overhead by this amount. The cost of each produced unit was then set at the charge

according to the outpatient fee schedule multiplied by the cost-to-charge ratio for each department. Information about department output was found in the hospital's annual report for 1993.

4.4.5.1.2 Physical and occupational therapy, social workers, clinical nutrition

For these departments, the cost per patient (social workers) or per consultation (physical and occupational therapy) was found by dividing the total costs after allocation of overhead by the output for 1993.

4.4.5.1.3 Laboratory of cardiology

Fees for all the services provided by this laboratory could not be found in the outpatient fee schedule. Instead, an investigation of actual costs of these services in a similar hospital was used in the same way as described above [155]. Fees for some services that could not be found in this investigation were set by clinical judgement after discussion with the head of the section of cardiology . The fee for percutaneous transluminal coronary angioplasty was set at the fee fixed by the central health authorities.

4.4.5.1.4 Laboratory of haematology

The total cost of this department was NOK 268,334 (USD 35,778), or only 0.3% of the total costs of the department of internal medicine. The services of this laboratory were not registered for each patient. Its costs were included in the department's overhead costs, and as such allocated to the wards according to number of admissions.

4.4.5.1.5 Laboratory of pulmonary physiology

Of the services provided by this laboratory, only bronchoscopies, which were the most costly, were registered for the individual patient. The cost of bronchoscopy was calculated as for the other service departments. Clinical judgement was used for setting the fees of some services which were not found in the outpatient fee schedule.

4.4.5.1.6 Referrals to other departments

Patients are sometimes referred for evaluation by physicians in other departments. We are not aware of any commonly accepted method for calculating the costs of such referrals, and they were not included in the present analysis.

4.4.5.2 Pharmaceuticals

The total cost of each pharmaceutical for the department of internal medicine in 1993 was obtained from the database of the hospital pharmacy. Drugs which accounted for more than 1% of the department's total drug costs were identified. The use of these drugs by each patient was registered from the medical records. The costs of un-registered drugs were allocated to each patient on the basis of length of

stay separately for each ward (see section 4.4.6.2). In 6 of the medical records, the drug prescription forms could not be found, but in all cases it was possible to infer the drugs used from information in the discharge reports.

4.4.5.3 Wards

4.4.5.3.1 Medical wards, including the coronary care unit

For each ward, the allocated overhead costs, the allocated physician labour costs and the nurse labour costs were divided by the total number of patient-days for 1993 to obtain the cost of one patient-day for each of these services. The cost of un-registered drugs for the individual patient was found by subtracting the cost of the registered drugs from the total drug costs of each ward (see previous section). The result was divided by the total number of patient-days for each ward to obtain the cost of the un-registered drugs per patient-day.
The ward costs not accounted for by the categories physician and nurse labour, overhead or drugs were labelled "hotel costs" and also divided by the number of patient-days to obtain unit costs.

4.4.5.3.2 Intensive care unit

The costs of this unit were treated in the same way as for the wards. Of the patient-days included in the study, only 10 were spent in the intensive care unit.

4.4.6 Calculation of cost of each hospital stay

The cost of each stay was calculated as the sum of the cost of resources registered for each patient, and the cost of resources apportioned to the patients on the basis of length of stay in each ward.

4.4.6.1 Costs of resources registered for each individual patient

For each stay, all diagnostic and therapeutic interventions were registered from the computerised and manual databases of the different service departments as well as from the medical records. The cost of each resource was calculated from the unit costs. A few resources were not registered:

4.4.6.1.1 Electrocardiograms

Electrocardiograms are routinely taken of all admitted patients. The cost of one was NOK 45 (USD 6). We included the cost of one electrocardiogram for each stay.

4.4.6.1.2 Sternal punctures and bone marrow biopsies

These tests were performed by the laboratory of haematology, see section 4.4.5.1.4.

4.4.6.1.3 Radiation therapy

Nine patients received radiation therapy in the department of oncology during their stays. The use of this treatment was not registered for practical reasons.

4.4.6.1.4 Pharmaceuticals

These were partly registered for each patient from the medical records, partly allocated according to length of stay, as described in sections 4.4.5.2 and 4.4.5.3.1.

4.4.6.2 Costs apportioned according to length of stay

Length of stay in each ward was obtained from the hospital database. The costs of nursing and physician labour, overhead, "hotel" and un-registered drugs were calculated separately for each ward, including the coronary care unit and the intensive care unit.

4.5 Statistical methods

The distributions of LEG, LQG and HYE were highly skewed to the left because of a high proportion of observations with the value zero. For this reason, statistical techniques making assumptions about normality of the distributions could not be used. Neither would any transformation make the distributions more normal because any transformed distribution would still have the same proportion of observation with identical values.

The bootstrap algorithm makes no assumption about the distribution of the observations and can be used for estimating confidence intervals in this situation [156,157]. The algorithm was implemented in a Microsoft Excel spreadsheet, using a random number generator for obtaining resamples and a simple macro for iteration. Software made especially for this purpose would have been much faster, but the use of a spreadsheet has the advantages of rapid implementation and easy debugging.

In multivariate linear regression analyses with one of the abovementioned variables as the dependent, inspection of the residuals made it clear that their variances were not constant, and that they were not normally distributed. For the same reason as above, transformations could not solve this problem. This precluded the use of ordinary methods for calculating confidence intervals for the regression coefficients which were therefore also estimated with the bootstrap algorithm.

Other statistical methods have been described in the individual papers.

5. Results

5.1 Agreement between the expert panels about assessments of health benefit and appropriateness

On inclusion, each admission was given a probability of 0.10 of being randomly assigned to group 1 for which discharge assessments were made by both expert panels for the purpose of studying inter-panel agreement (n=57) (Figure 2). The results for the assessments of LEG and LQG can be found in Paper 1. These assessments were classified in categories of no/low, intermediate and high gain. Agreement was measured with the weighted kappa statistics, which was 0.45 (95% confidence interval 0.18 to 0.73) for LEG and 0.63 (95% confidence interval 0.45 to 0.80) for LQG. This level of agreement is commonly characterised as "fair to good" [158]. It was only slightly lower than that found for other commonly used clinical methods [159], and higher than found in a review of agreement of peer assessment of implicit evaluation of patient-care episodes [160].

To investigate the ability of the panels to identify groups with either high or low gain, the agreement was also studied with a method based on log-linear models. This demonstrated better agreement about assessments in the highest and lowest categories for both measures.

Paper 1 also included a detailed description of the methods used for assessing LEG and LQG, as well as a discussion of methodological problems.

Paper 3 reported the results of the agreement study for the measures of short-term quality of life gain, i.e. the gains in quality of life below the sensitivity threshold of the time trade-off method. The weighted kappa statistic for HSQG was 0.70 (95% confidence interval 0.62 to 0.79) and for NHSQG 0.08 (95% confidence interval -0.20 to 0.35) (n=57). While there was no agreement about NHSQG, the kappa statistic for HSQG corresponded to good agreement.

Paper 5 included the result of the agreement study for the assessment of appropriateness. The overall agreement was 0.75, the kappa statistic 0.41 (95% confidence interval 0.15 to 0.68), i.e. fair agreement.

5.2 Health benefits from admissions to a department of internal medicine

Paper 2 and 3 reported the panels' assessments of the gains in life expectancy and quality of life for the patients randomised to group 2 and 3 (n=422)(Figure 2).

5.2.1 Gain in life expectancy (Paper 2)

The distribution of LEG was skewed to the left with 61% achieving practically no gain (<=0.10 years) while 5% had gains of 10 years of more (n=422). The mean LEG was 2.3 years (95% confidence interval 1.7 to 2.8). High age and the disease category "undiagnosed symptoms" predisposed for lower gain in a multivariate regression analysis, and "endocrinological disease" for high gain. Only one patient was judged to have experienced loss in life expectancy as a result of the stay.

A probabilistic sensitivity analysis was performed to study the possibility that negative effects had been under-estimated and positive effects over-estimated. A mean life expectancy gain of 1.4 years was found when assuming a rate of iatrogenic life expectancy loss 30 times that observed, with each case experiencing a loss corresponding to 50% of the average remaining lifetime of a person of the same age and sex in the general population.

5.2.2 Gain in quality of life (Paper 3)

LQG measured with the time trade-off method also had a left-skewed distribution. 59% had LQG equal to or less than 0.00, while 2% achieved gains >=0.50 (n=422). The 59% without LQG consisted of 40% with only HSQG and 19% with no health-related quality of life gain.

In a multivariate regression analysis with LQG as the dependent variable, high age, emergency admissions and the diagnostic categories "endocrinological diseases" and "pneumonia and influenza" were associated with higher gain ($P<0.05$). The categories "undiagnosed symptoms" and "cerebrovascular diseases" were associated with lower gain ($P<0.05$).

Since there was no agreement about NHSQG, it cannot be excluded that some of the admissions without health-related benefits may have had improvements in quality of life as an unspecific effect of hospitalisation.

5.2.3 Diagnostic gain for patients not experiencing health benefit

In addition to the health benefit assessments, the expert panels assessed whether the admissions had resulted in diagnostic gain (Appendix 1 and 2). The result of these assessments have not been reported elsewhere and are included here for the sake of completeness. The relation of this type of gain to health benefit was discussed in section 4.3.1.

The kappa statistic for the assessment of diagnostic gain in the agreement study was 0.57 (95% confidence interval 0.34 - 0.80), i.e. good agreement (n=57). Of the 72 patients without either LEG, LQG or HSQG in group 2 and 3, 38 had received a diagnosis that provided an explanation of disease manifestations (n=422).

5.3 The relationship between appropriateness and cost

In Paper 4, the LEG and LQG of group 2 and 3 (n=422) were expressed as gain in HYE (ΔHYE) (the more well-known term "QALY" was used in both Papers 4 and 5). The mean ΔHYE was 2.3 per admission, and its distribution is shown in Figure 3. Seventy-two (17%) admissions resulted in neither ΔHYE nor HSQG, i.e. in no health-related benefit. Thirty (7%) of the admissions with either type of benefit could have obtained the same benefit on a lower care level. Consequently, there were 102 (24%) inappropriate admissions.

The direct costs to the hospital from each stay were estimated. The inappropriate admissions had a lower mean cost (NOK 18,990 or USD 2,532) than the appropriate (NOK 43,500 or USD 5,800) (difference USD 3,268; 95% confidence interval 1,025 to 5,511). When adjusting for the effects of gender, age, admission category and diagnostic category in a multivariate regression analysis, appropriate admissions were still associated with higher costs ($P<0.001$). The 24% inappropriate admissions accounted for 12% of the total costs.

5.4 Prediction of appropriateness and potential for cost reductions

In Paper 5, the sensitivity and specificity for predicting that an admission would be appropriate were estimated with the discharge assessments as the gold standard (n=422). The potential costs saved and HYE lost from excluding the predicted inappropriate admissions were estimated. Elective and emergency admissions were analysed separately.

For elective admissions (n=152), the sensitivity was 89% and the specificity 31%. Denying admission for the 18% of elective admissions predicted to be inappropriate would have resulted in savings of 9% of the total costs (95% confidence interval 5% to 15%). At the same time, 5% of the ΔHYE from elective admissions would have been lost (95% confidence interval 1% to 12%). If the sensitivity and specificity had both been 100%, the number of elective admissions could have been reduced by 34%, and a cost reduction of 17% would have been achieved.

For emergency admissions (n=270), the sensitivity was 88% and the specificity 24%. Excluding the 14% inappropriate emergency admissions would have resulted in savings of 14% of the total costs (95% confidence interval 5% to 26%), and 18% of the total ΔHYE from emergency admissions would have been lost (95% confidence interval 6% to 34%). If the sensitivity and specificity had both been 100%, the number of emergency admissions could have been reduced by 19%, and a cost reduction of 10% would have been achieved.

If the predicted inappropriate admissions had been excluded, the savings per HYE lost would have been USD 3,910 (95% confidence interval 1,887 to 21,548) for elective admissions and USD 1,693 (95% confidence interval 474 to 6,525) for emergency admissions.

A multivariate regression analysis demonstrated differences in the predictions between men and women for elective admissions. If predicted inappropriate elective admissions had been excluded, a higher percentage of costs would have been saved (17% *vs.* 5%) and HYE lost (12% *vs.* 2%) for women than for men.

6. Discussion

6.1 Health benefit

6.1.1 Validity of the health benefit assessments

A measure is said to be valid if it is unbiased relative to a gold standard. In the present study, the gold standard for the health benefit assessments would have been the results obtained by randomising patients to admission or denial of admission. This would have provided a control group and made it possible to assess the effects of the hospital stays as such. However, though at least two studies in the 1970s did use this design for selected patients [161,162], practical and ethical problems made this approach unfeasible in the present study.

Another way of obtaining a control group would have been to look for patients who had for some reason been denied admission, but who were otherwise comparable to the admitted patients. This method has been used in a study comparing the mortality in an intensive care unit to the mortality of patients who had been refused admission because it was full or lacked trained nurses [163]. In the present study, it would have been impossible to obtain a control group of sufficient size, since virtually no-one is refused admission because of lack of capacity. Because of the heterogeneity of the study population, the number of included admissions would have had to be very high to ensure comparability between two groups for all relevant variables. Even in the referred study, there was a difference in case-mix between the two groups.

Accepting that it was not possible to obtain a control group in this study, the second best method would have been to assess health benefits on the basis of randomised controlled trials (RCTs) of the therapeutic interventions the patients had undergone. For several reasons, this was not possible. First, it is well known that many common interventions have not been evaluated with RCTs. Ellis et al estimated that only 53% of the treatments used in a department of general medicine were supported by this kind of evidence [164]; a finding that was later reproduced in a study by Michaud et al [165]. Second, though many RCTs use gain in life expectancy as an endpoint, few have so far included gain in quality of life. Third, there is a difference between demonstrating a treatment *efficacious* in the carefully controlled setting of a RCT and its *effectiveness* when used in daily clinical work [166]. Last, the patients included in

RCTs are often not representative of patients seen in clinical practice. In particular, patients with complicating diseases are commonly excluded, which makes it difficult to apply the results directly to all patients.

This left us with methods relying on clinical judgement, which have also been used in almost all other studies of inappropriate health care [43,59,71]. This does not mean that clinical judgement was used *as opposed to* a method based on RCTs. The experts were instructed to use the best available evidence in each case, preferably the results of RCTs. However, for cases where there was not sufficient empirical evidence for making an estimate, the experts had to use their clinical expertise.

The method involved separate assessments of the expected health of the patients with and without hospital admission, i.e. assessments of health in the future and in a hypothetical situation. For some conditions, these estimates can probably be made with a high degree of accuracy: A patient with meningococcal septicaemia who would otherwise have died, may be restored to full health after successful treatment. For other conditions, there will be greater uncertainty: An elderly overweight patient with diabetes mellitus and manifestations of generalised atherosclerosis discharged after having been treated with a thrombolytic agent for acute myocardial infarction, would probably have had a worse prognosis without this treatment, but how great would his life expectancy gain be? Comorbidity prevents us from applying the results of RCTs directly to such cases.

There is little reason to assume that this and similar assessments can be made with a high degree of precision and accuracy compared to the hypothetical gold standard discussed above [167]. However, we would expect the clinical experts to be able to make valid judgements on an ordinal scale of broad categories of gain. A similar assumption about physicians' predictive abilities underlies all ordinary clinical practice. Although health gain, taken as the difference between two assessments, has not been investigated, some studies have examined clinicians' ability to predict survival (Table 3). As expected, predictions of length of survival were not very accurate, but the most of the studies showed a positive correlation between predictions and actual survival. Several studies demonstrated good discriminative abilities for assessments of probability of short-term survival. In most of these studies, the estimates were made by only one physician. Estimates based on consensus methods would probably perform better.

Even if these studies lend some support to the assumption that panels of expert clinicians can make rough estimates of health benefit, we cannot deny the fact that the method used in this study has not been formally validated. The reasons for using it anyway was threefold:

- At present, there is no other method for estimating health benefit for unselected patients to departments of internal medicine. An increase in medical knowledge will hopefully enable us to make more precise estimates in the future. However, there will probably always remain a gap between the knowledge provided by research and that needed for assessment of the individual patient. This gap will have to be filled by clinical judgement, as also admitted by the proponents of so-called Evidence-Based Medicine [183]. Therefore, in studies like the present, we will most likely never be able to do without clinical judgement.
- Although assessment of health benefit was one of the aims of this study, its primary aim was to explore whether clinical judgement could in principle be used for identifying inappropriate admissions at the time of admission (section 3). At the present, it is difficult to imagine a method for classifying admissions according to appropriateness without relying on clinical judgement, either directly or for validating screening instruments. Consequently, assessing appropriateness on the basis of expert clinicians' estimates of health benefit seems justified.
- Estimates of the health benefits from health care programs are needed now. Important decisions about priority setting and resource allocation are made from surrogate measures of the effect of health care as number of treated patients, waiting list lengths and others. It is tacitly assumed that these measures correlate with the issue of real concern, i.e. health benefit. One good example is the use of utilisation review in the USA to reduce the number of hospital admissions without knowing how this affects the health of patients. It seems that an attempt to use the best, although imperfect, method we have to assess health benefit directly is worth the effort, and that it is one useful step forward relative to our present knowledge. A wise man has said: "Imperfect information at the point of decision is more useful than perfect information after it has been taken".

6.1.2 The magnitude and distribution of health benefits

When health benefits were estimated as ΔHYE and HSQG, the distribution of gains across the patients was left-skewed. Most patients were judged to have experienced small benefits while a few obtained gains corresponding to life-saving treatment. The gain of these few was so great as to result in a mean ΔHYE of 2.3. Because most of the ΔHYE was won as LEG, and little as LQG alone, this result was similar to the mean LEG.

The mean ΔHYE may seem high compared to other interventions. Wright and Weinstein recently tabulated the gains in life expectancy from a variety of medical interventions (Table 4) for use as a benchmark for new interventions. The table was compiled from publications found by searching MEDLINE [184]. The mean LEG found in the present study, 27.6 months, was higher than most treatments on the list. However, information about the effects of many of the more effective interventions used in internal medicine cannot be found in MEDLINE, simply because they have never been subjected to controlled trials. Antibiotics for serious infections, hormone substitution for failure of endocrinological organs and hemodialysis for renal failure are examples of treatments that are considered to be life-saving, and which for ethical reasons will never be subjected to RCTs. Table 2 in Paper 2 shows that several of the patients with the highest gains in the present study had received such treatments. One third of the total LEG was achieved by these ten patients. The patient with the highest gain was an 18 year old woman who was treated for meningococcal septicaemia. The panel judged the treatment to have been life-saving and to have gained 63 life years.

There is, however, little reason to believe that the experts have been able to make an accurate estimate of the mean health benefit measured in HYE. Even so, all kinds of bias that do not affect the rank order of gain would have resulted in the same shape of the distribution curve. More confidence should therefore be placed in this than in the numerical estimates. It is probably characteristic of departments of internal medicine that the benefits of most patients are small, but that the life of an occasional patient is saved. I am not aware of other studies of health benefit from consecutive hospital admissions. It would be interesting to know whether the results in e.g. a surgical department would be different.

HYE and QALY have been developed for measuring and comparing health gains. Most of the studies using these measures focus on chronic conditions with health states of stable reduced quality of life. It is significant that 28% (Paper 4, Table 1) of the admitted patients had gains that were below the sensitivity threshold of the instrument we used for measuring ΔHYE. The reason was that the time scales of the instrument could not register the very small trade-offs made by relatively healthy persons with temporary reductions in quality of life. For instance, a 20 year old man with painful tonsillitis from infectious mononucleosis would probably be willing to trade off very little of his expected lifetime of 53 years to obtain full health for the 7 to 10 days this condition lasts. If we assume a quality of life of 0.3 for 10 days, the maximum gain from the palliative treatment available for this condition would be 0.02 HYE, which could not have been registered by our time trade-off instrument.

Though time trade-off instruments for measuring temporary reductions in quality of life exist, I know of no study where it has been used for measuring health benefits as small as this. Comparing the hypothetical mononucleosis patient to the patient with the highest gain in our study (63 HYE), there is a difference in gain of 3 orders of magnitude.

It is important that treatment of conditions of short duration will result in small gains relative to the high-gainers no matter how severely their quality of life is reduced. Admitting these patients for hospital treatment is now commonly accepted, but if rationing on the basis of cost-effectiveness should be implemented, these treatments would have to have very low costs to defend their place. This seems to be in conflict with our instinct that acutely ill patients should be treated, and in practical life, other factors than cost-effectiveness alone may play a role when deciding whether resources should be allocated to such conditions.

6.1.3 Factors associated with health benefit

In the search for factors predisposing for health benefit, multivariate linear regression analyses with LEG and LQG as dependent, and age, gender, admission category and diagnostic category as independent variables, were performed (Paper 2, Table 3, and Paper 3, Table 3). As expected, the diagnostic categories were the most important regressors. Since the diagnostic categories in the analyses were heterogeneous, and comorbidity was not taken into account, little weight should be

attached to the actual parameter estimates except to demonstrate that diagnosis matters.

Age was negatively associated with LEG, but the regression coefficient was low compared to those of the diagnostic categories. An increase in age by 10 years only decreased the expected LEG by 8 months. The presence of an endocrinological disease increased the expected LEG by 9.9 years, and undiagnosed symptoms decreased it by 2.1 years relative to the reference "other". Accordingly, as judged by the expert panels, age was not an important determinant of LEG. However, cases can easily be imagined where age must play a more important role for LEG. The 18 year old patient who was cured for septicaemia no doubt would have had a lower LEG if she had been 50. On the other hand, for some conditions, e.g. a non-curable malignant disease, the disease itself sets the limit for the LEG obtainable by treatment that has no chance of restoring the patient to full health. This is more typical of chronic diseases. Presumably, the experts considered most of the cases in the study to fall in this category. It is, of course, also possible that they under-estimated the magnitude of the negative association between age and LEG.

Age was positively associated with LQG. Again, the regression coefficient was low compared to those of the diagnostic categories. An increase in age of 10 years increased LQG by only 0.01, while the presence of endocrinological disease increased it by 0.15.

The regression analyses of LEG and LQG resulted in adjusted R^2 of 0.17 and 0.13 respectively, meaning that little of the variance of these two measures could be explained by the variables in the model. The low precision of our method for measuring health benefits was also an important source of variation. A more detailed classification of diagnosis and disease severity might have improved the fit of the model.

6.1.4 Admissions resulting in no health benefit

This study was designed to detect all direct benefits to the admitted patients. The results of the agreement study demonstrated that the panels made reliable assessments about health-related benefits, but that the agreement about non-health-related short-term quality of life gain (NHSQG) was poor. Thus, some of the patients without health-related benefits may have had benefits from hospitalisation that were

not caused by specific care or treatment. However, of the 72 patients without health-related gain, none were scored in the category "high" and only 2 in the category "intermediate" NHSQG. Even allowing for disagreement between the panels, these 72 patients (17%) probably had either no or very little benefit.

In the instructions to the experts, care of the dying was especially mentioned as a type of benefit that might qualify as NHSQG. However, half of the 20 patients who died in the hospital were scored for health-related benefits. This probably means that they received palliative treatment.

The possible reasons why a patient did not achieve health benefit from a hospital stay are listed in Table 5. The admissions were not classified according to these categories in the present study. In other studies, significant rates of inappropriate interventions have been found (see section 2.4). Many patients would also be expected to fall in the category of no available effective therapy.

6.1.5 Iatrogenic health losses

In studies from the USA, adverse drug events have been found in 2.4% [185] and 6.5% [186], and adverse events in general in 3.7% [187] of hospital admissions. One study reported that iatrogenic disease was the cause of 5.4% of hospital admissions [188], and another iatrogenic illness in as many as 36% of the patients admitted to a general medical service [189]. The percentages for some patient groups, e.g. the elderly, may be even higher [190]. There is little doubt that many of the events have serious consequences. In the Harvard Medical Practice Study, which reviewed 30,121 patient records, 13.6% of the adverse events led to death [187]. In a study from a department of internal medicine in Norway, adverse drug events were the probable cause of 11.8% of the in-hospital deaths [191]. Two other Norwegian studies have found percentages of adverse events as causes of admission to departments of internal medicine of 5 [192] and 7 [193], i.e. of the same magnitude as in the USA.

Health benefits from hospital stays should be balanced against the health losses resulting from adverse events. The health losses detected by the instrument used in the present study were small. Only 1 admission (0.2%) resulted in a negative LEG of 0.07 years, and only 3 (0.7%) in negative LQG. Although not strictly comparable, these percentages were lower than in the studies referred to above. The reason may

be that most adverse events are minor and have no consequences for patient health in the long-term. Also, the experts may have been biased towards over-estimating positive and under-estimating negative treatment effects.

Paper 2 reports a sensitivity analysis of assuming a much higher rate of negative LEG than observed. Based on data from other studies of adverse events, it was concluded that the total positive LEG by far outweighed the negative. As LEG generated most of the ΔHYE, it follows that the total ΔHYE was also positive by a wide margin.

The study did not include an instrument for measuring short-term quality of life losses from adverse events analogous to HSQG for positive gains. Since many treatments have high rates of adverse effects, it may be safely assumed that many patients suffered short-term reductions in quality of life. The important question is whether these losses outweighed the benefits for some of the 28% who were judged to have had only HSQG (Paper 4, Table 1), and, consequently, whether this percentage was over-estimated. If so, the percentage of admissions without health benefits may have been under-estimated.

6.1.6 Care level

If a patient did achieve health benefit, the experts considered whether the same benefit could have been obtained in primary care or in the outpatient clinic. Accordingly, the assessments of necessary care level depended on the assessments of health benefit, and the kappa statistic for the agreement between the panels about this measure would incorporate the disagreement about both health benefit and care level. This was solved by estimating kappa for the combined judgement that the patient either had experienced no ΔHYE or HSQG, or that the same gain could have been achieved on a lower care level. This corresponds directly to the definition of an inappropriate admission used in this study (section 3). Fair agreement was found for this assessment (kappa 0.41).

For one of the benefit measures (NHSQG), the kappa statistic indicated no agreement between the panels. Though this measure was excluded from further analysis, disagreement about this measure could possibly have influenced the care level assessments and thereby lowered the kappa statistic for the combined judgement considered above. This could have happened if a panel had made the

judgement that, for some patients, hospitalisation was necessary for NHSQG, but not for ΔHYE or HSQG. However, this combination of assessments seems very unlikely, and it is a reasonable assumption that the disagreement about NHSQG did not influence the agreement about appropriateness to any noticeable extent.

The panels judged 24% of the admissions to have been inappropriate (17% had had no health benefit and 7% could have been treated on a lower care level). In comparison, studies from the USA have found percentages of inappropriate hospital admissions ranging from 10 to 40 with most observations in the interval 10 to 20 (see section 2.6). Our percentage was slightly higher. It was also higher than the results in most European studies, where percentages from 1.3 to 25 were found. However, the percentage was lower than in the two investigations from Kirkenes Hospital (35 and 48%) [87,89].

6.2 Cost

6.2.1 Appropriateness and cost

An important finding of the present study was that the inappropriate admissions had a lower mean cost than the appropriate. In a sensitivity analysis, this result was robust to variations in the estimates of unit cost. A multivariate linear regression analysis indicated that it was also independent of gender, age, admission category and diagnostic category. The regression was repeated without three appropriate admissions with lengths of stay of more than 6 months and very high costs. In this analysis, the regression coefficient for the appropriateness-variable was virtually unchanged and its P-value 0.0001. Repeating the analysis after changing the status of these three admissions from appropriate to inappropriate also gave essentially the same result.

Since the expert panels had information about the use of ancillary resources and length of stay when assessing appropriateness, the possibility of information bias from differential misclassification of appropriateness relative to cost must be considered. The expert panels might have been biased towards classifying stays using more resources as appropriate. This possibility was discussed in Paper 4, where it was concluded that the observed difference in mean cost was too large to be explained by this type of bias.

6.2.2 Limitations of the cost analysis

In the cost analysis, the use of resources was registered for each individual patient. Because it was impossible to perform on-site registration for this large number of patients, we could only register use of resources from the medical records and the hospital databases. "Hotel", and physician and nurse labour costs had to be allocated on the basis of length of stay, except for the labour costs included in the unit costs of ancillary services as e.g. gastroscopy or hemodialysis.

Since nursing was the second most important cost, the error made by this method could be substantial. Obviously, the use of this resource varies greatly across patients and should preferably have been registered for each individual. It was considered whether to use the nursing costs of the patients' DRG as basis for allocating these costs, but in the Norwegian version of the DRG-system, nursing costs have been based on clinical judgement and not on an investigation of actual use by patients [194]. In the present study, nursing costs were allocated according to length of stay at ward level. Using ward instead of department level assures at least some homogeneity in the patient group with regard to the need for nursing. This is especially important for the intensive and coronary care units, where nursing costs are high. It has also been estimated that a high percentage of nursing time is used for administrative tasks, which also provides some justification for this approach [155].

6.2.3 Marginal vs. average costs

For estimating savings from denying admission, marginal costs, i.e. the cost of treating one additional patient, would have been preferable average costs.. However, marginal costs depend on the time perspective of the cost analysis.

In the very short run, the total costs of personnel and major equipment would have been constant and the marginal cost would have been the cost of non-reusable resources consumed by the individual patient as laundry, food, drugs, intravenous fluids etc. For the hospital in general, personnel costs accounted for about two thirds of the total costs, and, consequently, in this time perspective, marginal costs would have been less than one third of the average costs. This assumes that the department operated below full capacity. If the capacity was exceeded, extra labour would have had to be bought at a high cost, and marginal costs might have been

higher than the average. As referred in Paper 5, the median bed-occupancy rate in the study period was 0.84, which does not indicate high costs for extra labour. In the long run, costs can be saved by downscaling the entire hospital, which means that all costs would decrease proportionally. In this situation, the savings obtained would approach our estimates of average costs.

6.2.4 The time perspective of the cost analysis

The cost analysis included only direct costs to the hospital incurred during the included stays. For patients with subacute or chronic conditions, interventions planned during these stays would sometimes be performed in the outpatient clinic or in another department after discharge. Thus, the decision to hospitalise the patient in the first place could cause additional direct costs to the hospital after discharge. The inclusion of costs only during the included stays could under-estimate the cost reductions that would follow from a decision to deny admission.

An alternative method would have been to include the costs of all interventions planned during the stay. However, it is often difficult to decide what has been explicitly planned. Some patients are discharged after having been scheduled for e.g. a coronary artery bypass graft or a percutaneous transluminal angioplasty, the costs of which would have been included with this method. Others are discharged with plans for performing further investigations, e.g. myocardial scintigraphy, that may ultimately lead to the decision to perform these procedures. In these cases, the cost of a revascularisation would not have been included. The same problem would apply to many other patients who are discharged with appointments for follow-up.

Some of the difficulties could be solved if one chose a method that included all direct costs to the hospital for a set time period. However, the problem of the cost of lifelong chronic conditions in need for continuos follow-up remains. It is hard to see how this would be an improvement over the method chosen, i.e. to restrict the cost analysis to the included stay.

It could be argued that the limited time perspective of the cost analysis prevents us from drawing conclusions about the relationship between appropriateness and cost. More specifically, it could be that an inappropriately admitted patient with low costs during the included stay was scheduled for costly interventions later, which would tend to invalidate the conclusion that inappropriate admissions are less costly. To

investigate this problem without performing an actual analysis of costs after discharge, we registered plans for follow-up and interventions for each stay. Fewer appointments were made for inappropriately admitted patients (Table 6)($P<0.001$). In particular, fewer of them were scheduled for surgery (1 vs. 8%), which would be expected to incur the highest costs. Although it does not constitute a proof, this suggests that the these admissions would also have had lower costs in a longer perspective.

While the cost of each stay can be estimated, there is no analogous method for estimating the health benefit that results from a single hospital stay in a series of stays and other interventions, as discussed in section 4.3.2. For this reason, we defined the health benefit from an included stay as the benefit *for which it was a necessary condition.* This method introduced an incongruence between the methods for estimating costs and benefits. As explained in the previous paragraph, it was not possible to define a series of future contacts with the hospital for which costs could be estimated, as would have been preferable. Accordingly, the cost-effectiveness of the admissions could not be estimated, although some considerations of the upper bounds of the cost-effectiveness-ratios relative to that of other interventions are presented in a later section.

6.3 Prediction of appropriateness

6.3.1 Sensitivity and specificity

At admission, the panels were generally too optimistic about the results of the hospital stays. The number of inappropriate admissions was predicted to be 66 (16%), whereas 102 (24%) was observed (Paper 5, Table 1).

In clinical epidemiology, sensitivity and specificity are indices used for characterising a diagnostic test. The results of the diagnostic test are compared with a gold standard and the indices calculated according to standard formulae [195]. In the present study, the panels' predictions of appropriateness at admission can be thought of as a diagnostic test, and the appropriateness as judged by the other panel after discharge as the gold standard.

For the prediction that an admission would be appropriate, a sensitivity of 88% and a specificity of 27% were found. Thus, while the majority of appropriate admissions can be identified at admission, the panels' ability to detect the inappropriate was poor.

One might assume that the reason for this could be that the threshold for judging that anyone had had health benefit had been set very low, and that it would have been easier to predict very high gains. When the specificity for predicting gains >= 10HYE was calculated, a higher value was found (96%), meaning that the experts were able to identify almost all patients with lower gains. However, this would have been achieved at the cost of a lower sensitivity (33%), meaning that two thirds of the patients achieving these high gains would not have been identified.

For elective admissions, the panels had lower sensitivity and higher specificity for women than for men. This finding was discussed in Paper 5.

6.3.2 Possible causes of poor predictions

6.3.2.1 Lack of direct contact with the patients

The validity of the medical record as basis for assessment of quality and outcome of care has been questioned in some studies. Fessel et al found considerable disparity in the frequency of documentation of common symptoms of appendicitis in three different hospitals, but no association between documentation and diagnostic accuracy or outcome [196]. Romm et al reported incomplete recording of information when comparing the medical record to transcripts of outpatient visits [197]. Burns et al found paucity of information about functional status in the medical record compared with patient self-report [198].

On the other hand, the medical record has been found sufficient for several purposes, including detecting adverse events [199,200], finding the indication of medical procedures [201] or judging impairment of organ function [202]. One study examined the influence of the completeness of the medical record on identification of inappropriate days of care with the Appropriateness Evaluation Protocol. Significantly higher rates of inappropriateness were found for lower levels of completeness. However, the differences were small and not significant when adjusted for other factors associated with inappropriateness [203].

If predictions of health benefit and care level should be used for reducing hospital admissions, they would have to be made by the admitting physicians, who would have the advantage of direct contact with the patient. If this should enable better predictions than the expert panels, these would have to rely on information that was not documented in the medical record or communicable in written form. In the present study, the first possibility was counteracted by letting the project co-ordinator check the medical records for missing information the day after admission. Even if the protocol did not allow him to obtain additional information directly from the patients, this procedure would seem to exclude the possibility that information systematically omitted from the medical records was responsible for poor predictions.

The second possibility would imply that some form of global assessment of the patient was an important factor for predicting health benefit. At least one study has found that clinicians agree poorly on such assessments [204]. Even if such factors did play a role, the experts had several advantages relative to the admitting physicians. First, they had longer experience than the average intern or resident. Second, they had the opportunity of discussing difficult cases with the equally experienced members of the panel. Last, they probably had more time for considering each problem. In all, there is little reason to believe that the admitting physicians would have made better predictions than the panels.

6.3.2.2 The composition of the expert panels

Instead of using three internists, specialists from two related disciplines were chosen. A similar design was used in a recent study of the health benefit from helicopter evacuation [152]. Some studies have shown that the composition of expert panels matters for their assessments, but it is not known which composition is optimal. Leape et al found that, for carotid endarterectomy, a panel composed exclusively of surgeons found fewer operations inappropriate than a multi-disciplinary panel [205]. Similar results were reported by Scott et al for cholecystectomy [206]. Coast et al used two panels consisting of general practitioners and one consisting of consultants for assessing necessary care level. The consultants judged hospital care necessary for a higher percentage of the admissions than did the general practitioners [207]. Ayanian et al found that cardiologists rated coronary angiography as more appropriate than primary care physicians for some indications [208].

Whether these results reflect a tendency towards over-estimating the effects of one's own speciality, is not known. In the present study, one of the reasons for choosing different specialists was to eliminate this source of bias, if present among internists. However, we cannot exclude the possibility that panels consisting of three internists would have made better predictions.

6.3.2.3 The instrument

The present study used an instrument that was designed to consider all the criteria necessary for deciding whether an admission would result in health benefit on an appropriate care level or not. The instrument was rather complicated, and training and a detailed instruction manual were necessary. Even if a pilot study was performed to identify problems, the complexity of the instrument may have contributed to the poor predictions. Although this cannot be excluded, the agreement found for the instrument was comparable to that of other studies of peer assessment of patient-care episodes [160]. This makes it less likely that we could have obtained better predictions with another instrument.

Some factors may have made the assessment of care level difficult. The alternatives were specified as primary care, outpatient clinic or hospital admission. While the facilities in the outpatient clinic are well defined and well known by the experts, this may have been different for primary care. Some primary care centres include a general practitioner hospital, and other facilities differ as well. Since the summaries were blinded with respect to geographical data, the experts had to rely on their concept of the facilities typically available. Better predictions of care level might have resulted from a better specification of the alternatives to hospital admission.

6.3.2.4 Disagreement about health benefit assessments

Two expert panels were used to avoid bias from letting one expert panel assess the same admissions both at admission and after discharge. Otherwise, the predictive abilities might have been over-estimated because the panels could have remembered their predictions at admission when making their discharge assessments. Another consequence of this design was that the quality of the predictions also reflected the inter-panel variation in estimation of health benefits. Even if the panels had been able to predict their own discharge assessments

perfectly, their predictions of the other panel's assessments might have been poor if the inter-panel variation was high.

To examine this possibility, the study included an investigation of the agreement between the panels (section 5.1). The agreement was found satisfactory and comparable to the agreement about other clinical methods [159]. Even so, we would expect the disagreement between the panels to have made some contribution to the poor quality of the predictions. To assess the magnitude of this effect, one panel's sensitivity and specificity for "predicting" the other panel's estimates of the same cases in the agreement study were calculated (patient group 1 in Figure 2). A sensitivity of 83% and specificity of 59% were found (n=57). Since the corresponding values for the predictions made at admission were 88% and 27% (n=422), it was easier for the panels to "predict" which admissions would be judged inappropriate by the other panel when the information collected *during* the stay was available. However, the rather low specificity (59%) demonstrates that the disagreement between the panels also made a contribution to reducing the quality of the predictions at admission. Variation in the assessments of inappropriateness between admitting physicians would also occur in clinical practice. This constitutes an additional difficulty when trying to use clinical judgement for reducing inappropriate admissions.

6.3.2.5 Clinical uncertainty

Clinical uncertainty at admission will always prevent us from attaining perfect sensitivity and specificity when predicting appropriateness. Some patients are admitted without a conclusive diagnosis while there is uncertainty about the effect of planned treatment for others. In these cases, there is not sufficient information for making an accurate prediction. This was probably the most important cause of the poor predictions. In daily clinical work, most clinicians deal with this uncertainty by keeping the threshold for admission low.

Because emergency admissions are not planned, very little can be done at the time of admission to reduce the uncertainty by obtaining more information. The situation is different for elective admissions, of which 34% (Paper 4, Table 4) were inappropriate. By communicating with the referring physician, the doctor in charge of planning the admission has the opportunity to let the patient undergo diagnostic

interventions in primary care or an outpatient clinic before admission. The possibility that this could improve the predictions should be investigated.

6.3.2.6 Spectrum bias

The mediocre ability of doctors to predict which patients would benefit from a hospital stay is worrying since a reliable gate-keeper function is crucial for our health care system. However, the sensitivity and specificity were estimated for the population of patients actually admitted to the department. This population was already highly selected by the referring physicians, and, accordingly, different from the population of all patients who might have been candidates for admission. The primary care physicians had already sorted out most of the patients that could easily have been identified as inappropriate, and the studied population may have consisted of patients with an obvious need for hospitalisation in addition to a small number of "problem" cases. The result was a population that generated low specificity for predicting appropriateness by so-called *spectrum bias* [209].

If we take primary care physicians' referral of a patient for admission to represent their prediction that the admission would be appropriate, there is good reason to assume that their specificity for detecting appropriateness must have been higher than that of the expert panels. Since only a minority of all consultations in primary care results in hospital admission (for emergency consultations about 10% [210]), only a few of the potentially inappropriately admitted patients are actually admitted (unless the percentage of inappropriate patients in the population is very low, which is unlikely). This means that the primary care physicians' specificity for detecting and admitting appropriate patients in this population was probably better than that of the expert panels' for the actually admitted patients. Of course, we know nothing about the sensitivity, which might be poor, i.e. that patients who would have benefited from a hospital stay may not have been admitted. However, the decision to admit a patient or not is seldom final. By using time and the course of the disease as diagnostic tools, general practitioners may also attain a reasonable sensitivity for detecting those patients who will benefit from a hospital stay.

6.3.2.7 *Bias in the final assessments of appropriateness after discharge*

In section 6.1.1, the validity of the health benefit assessments made after discharge was discussed at length. Since the quality of the predictions of appropriateness were judged by comparison with these assessments, it should be considered how their validity might have affected the conclusion that this quality was poor. Suppose that a method which had formally been demonstrated as valid had been used for the discharge assessments, and that the actual discharge assessments of the panels were poor compared to this method. Could the sensitivity and specificity of the predictions have been higher if they had been compared to this hypothetical gold standard? Since this assumes that the predictions could have been better estimates of the gold standard than the panels' discharge assessments, the answer is no. At discharge, the experts used the same instrument and had access to the same information that was available at admission and, in addition, all information gathered during the stay. If it is assumed that more information must lead to better estimates, the discharge assessments must have been better estimates of the hypothetical gold standard than the predictions. Accordingly, the predictions would also have been poor compared to this gold standard.

This has the important consequence that the conclusion that the predictions were poor did not depend on the validity of the panels' discharge assessments.

6.4 The potential for cost reductions

6.4.1 Cost reductions and health losses

The purpose of trying to predict appropriateness was to explore the potential for cost savings and health losses. Elective and emergency admissions were analysed separately. For both types of admission, modest cost reductions could have been obtained (9 and 14%), but at the cost of a loss in HYE for patients falsely predicted to have been inappropriately admitted (5 and 18%)(Paper 5, Table 5). The savings per HYE lost would have been NOK 29,328 (USD 3,910) for elective admissions and NOK 12,699 (USD 1,693) for emergency admissions. The main reason for the difference was that more HYE would have been lost per patient for emergency admissions. Repeating the analysis after excluding the 3 outliers mentioned in section 6.2.1 lowered the savings for emergency admissions to NOK 6,561 (USD

875) per HYE lost, but gave the same result for elective admissions. If these 3 admissions had all been predicted to be inappropriate, the savings would have been NOK 73,187 (USD 9,758) per HYE for elective and NOK 17,432 (USD 2,324) for emergency admissions.

Costs saved per HYE lost from not performing an intervention, in this case admitting a patient, is equivalent to the costs expended per HYE won from performing it. In a recent study, Nord et al discussed the cost per QALY of different treatments [211]. Only hip replacement had a lower cost per QALY (NOK 12,750 or USD 1,700), whereas the cost per QALY of 9 other referred treatments ranged from NOK 42,750 (USD 5,700) to NOK 727,500 (USD 97,000). A commonly cited upper limit for cost-effective care has been NOK 375,000 (USD 50,000) per QALY [212,213]. Consequently, the relationship between costs and health losses of the proposed strategy for reducing admissions would have been less favourable than for many other interventions commonly accepted as cost-effective, even if there seems to be little theoretical support for the USD 50,000-limit. Additional support for this conclusion can be found in a study of life-saving interventions by Tengs et al. They found that the median cost per life-year won of 310 medical interventions was USD 19,000 [214]. Considering that the cost per HYE would have been higher, the cost of the HYE won by not trying to reduce the number of admissions in our study would seem low.

In section 6.3.2.1, the probability that the admitting physicians would have made better predictions than the panels was discussed. Paper 5 examined the effects of improved predictions in a sensitivity analysis. Even with a sensitivity of 90% and specificity of 50%, equal costs of the inappropriate and appropriate admissions, and an over-estimation of HYE by 100%, the cost saved per HYE lost was still only NOK 195,984 (USD 26,131). This leaves considerable room for under-estimation of costs before the limit of NOK 375,000 (USD 50,000) per QALY could have been reached.

If we had estimated marginal costs in the short run instead of average costs, this ratio would have been even lower and our conclusion strengthened. The same would have been the case if it had been possible to estimate costs from a societal perspective, because the costs of treating some patients on a lower care level would have had to be subtracted from the estimate of savings in hospital. This assumes that the health benefit they would have achieved outside hospital would have been

negligible compared to what they would have obtained if admitted. Since the need for hospitalisation was explicitly assessed by the expert panels, this seems reasonable. Another possibility is that hospital treatment might have induced additional societal costs after discharge, which would have increased the savings from not admitting some patients. We considered this effect to have been small.

If a system for reducing the number of admissions on the basis of judgements of appropriateness had been implemented, these judgements would have had to be performed by the admitting physicians. Since the judgements would have been based on the same information that is routinely collected in today's system, the additional costs incurred would have been negligible and were not included in the cost analysis. If they had been, they would have had to be deducted from the savings, which would have strengthened our conclusion.

6.4.2 Future costs

In section 6.2.4, the problem of costs to the hospital incurred after discharge from the included stays, was discussed. The savings per HYE lost estimated in the previous section could have been underestimated because of costs of planned interventions and follow-up.

To explore this possibility, plans for follow-up of the 66 patients that would have been denied admission were registered from the medical records (Table 7). Such plans existed for 22 of them. The highest costs would probably have incurred for the 3 who were scheduled for readmission and for the 2 who were scheduled for surgery.

The potential savings form denying care for these 66 patients would have been NOK 1,972,389 (USD 262,985) and the potential health losses about 135 HYE (calculated form Table 4, Paper 4 and Table 5, Paper 5). If the NOK 375,000 (USD 50,000) per HYE-limit is tentatively accepted, the savings would have had to be NOK 50,505,013 (USD 6,734,002), meaning that the future costs for the 22 patients with further plans would have had to exceed NOK 48,000,000 (USD 6,400,000), which seems unlikely. Accordingly, a cost analysis including costs after discharge would probably not have changed our conclusion.

6.4.3 Care level

In section 6.3.2.3, it was discussed whether improvements could have been made to the instrument for predicting necessary care level. However, patients who could have been treated at a lower care level accounted for only 5% of the total costs, which indicates that the potential for additional savings by predicting this group perfectly was small.

Coast et al examined the alternatives to hospital care for acute admissions to a department of general medicine and care of the elderly in the UK [207]. Using a detailed list of alternatives and assessment by expert panels, she found that a lower care level would have been a possibility for between 5.5 and 14% of the patients. Her expert panels made their assessments on the basis of information available at admission, but only assessed patients found to have been inappropriately admitted by the screening tool ISD-A [71]. In another study, Coast et al made a cost analysis of the alternatives to hospital care and found that few resources would have been saved if these had been used [215]. Even if there are important differences between Norwegian and British health care, these results suggest that the potential for additional savings in our study from making a more detailed specification of the alternatives to hospital care, might have been limited.

However, this refers to the way primary and hospital care are organised at present. There are indicators that some of the treatments now reserved for inpatients could be used on lower care levels without health loss. One example is the treatment of myocardial infarction with streptokinase in primary care [216]. A restructuring of the care levels could permit more treatment outside hospital. Whether this would be more cost-effective than hospital care would have to be investigated.

6.4.4 Other studies of cost and predicted health benefit

I have not been able to find other studies of the relationship between predicted health benefit and costs from departments of internal medicine. However, Pompei et al studied charges and prognosis for 549 patients admitted to the medical service at the New York Hospital during a 1-month period in 1984 [96]. The 5-year prognosis was estimated as favourable or unfavourable by admitting residents. In contrast to our study, these estimates were made within 24 hours of admission, and consequently, some results of tests done after admission must have been available.

When comparing the estimates with mortality at one year, this was 9% in the "favourable" and 50% in the "unfavourable" category (P<0.01). Large expenditures were associated with patients who died in the hospital, especially those whose death was unexpected. Pompei concluded that the imprecision of clinical judgement at the time of admission in predicting long-term outcome argues for aggressive management of acutely hospitalised patients when there is any doubt about their prognosis.

Detsky et al let house officers estimate the probability of survival until discharge for 1,831 patients admitted to an intensive care unit. Although this study population was quite different from that in the present study, it will be reviewed briefly because of the similarities in design. Detsky found that expenditures were positively correlated with estimated probability of survival for non-survivors, but negatively for survivors. In other words, the highest expenditures were found for the patients with unexpected outcome. He concluded that prognostic uncertainty was important in determining resource expenditures for the critically ill [98]. Calculations from the data of Table 2 in Detsky's paper show that 6% of the expenditures had been saved at the cost of losing 1% of the survivors if he had chosen to admit only patients with a probability of survival of greater than 20%. Cut-off levels at 40 and 80% would have yielded savings of 14 and 46%, and loss of survivors of 3 and 25%, respectively. As it would probably not have been acceptable to deny admission for a patient with a probability of survival even as low as 20%, it would not have been possible to obtain savings based on prognostic assessments in this setting.

Although none of these two studies are directly comparable to the present study, their conclusions resemble ours in their emphasis on prognostic uncertainty as an important determinant of resource utilisation. This uncertainty causes clinicians to keep a low threshold for admitting patients. The threshold may be lowered further as a result of the decreasing tolerance of mass-media and the public for physician "malpractice".

6.5 Strategies for reducing length of stay

Another possibility for obtaining savings would have been to reduce the use of resources after admission. This could have been achieved by reducing length of stay, which was the most important determinant of cost. Observation units, where

patients undergo a rapid diagnostic work-up without actually being admitted, have been proposed as a way to quickly reach a decision about whether admission is necessary and to reduce costs [217,218]. One study found that the cost of asthma patients treated in an observation unit was lower than for admitted patients (USD1,202 versus USD 2,247) [219]. Another study examined the predictive abilities of physicians in an observation unit for detecting the presence of pathology necessitating hospitalisation for selected diagnoses. A sensitivity of 100% and specificity of 86% were found [220], raising the question of whether prediction of inappropriateness in general would also have been better.

In the present study, inappropriately admitted patients already had a shorter mean length of stay than others (4.3 vs. 10.0 days). The percentage of such patients was 24, and they used 12% of the resources. Let us assume that this group could have been identified with a sensitivity and specificity of 100% after one day in the hospital. If their use of resources was proportional to length of stay, we would have saved 12%*(4.3-1)/4.3 of the resources, i.e. 9%. But since, in reality, the identification would not have been perfect and more ancillary resources would probably have been used the first day, the savings would have had to be lower, and there would still have had to be some health loss. Consequently, the savings from reducing the length of stays resulting from inappropriate admissions would have been modest in our setting.

This leaves us with the possibility that the length of stay for appropriately admitted patients could have been reduced without reducing health benefits, which was not investigated in the present study. However, there is little doubt that some of the included patients stayed in the hospital longer than necessary. Three of the patients with benefits had stays lasting more than 6 months because of insufficient nursing home capacity. These 3 patients accounted for 12% of the total costs. Two of them achieved gains of 1 and 4 HYE, and one only low degree HSQG. At admission, it was erroneously predicted that the patient with the highest gain would not have benefited. If these 3 patients could have been discharged when their medical treatment was complete, the savings would probably have been of the same magnitude as when identifying inappropriate admissions after one day's stay. It must be assumed that the length of stay could have been reduced for other patients as well. However, there is probably a limit to the reduction in average length of stay that

can be achieved without increasing the per diem cost of nursing. In the USA, one study found that the amount of nursing per patient per day increased when length of stay was reduced [221].

Some studies have investigated interventions to reduce length of stay. In the USA, Wachter et al studied length of stay, cost, 6 month mortality rate, readmissions and patient functional status after reorganising half of the wards in an academic medical service to involve faculty members more in inpatient care. The other half was left unchanged. The hypothesis was that more expertise would reduce costs. When the wards were compared, mean length of stay was shorter (4.3 vs. 4.9 days; P=0.01) and mean cost lower (USD 7,007 vs. 7,777; P=0.05) for the reorganised wards. However, the cost difference, which was of borderline statistical significance, is difficult to interpret because the cost analysis did not include physician costs. It must be assumed that use of higher expertise in direct patient care would incur extra costs. There were no differences in patient outcomes [222].

Two recent British studies compared hospital at home care to ordinary inpatient care. Hospital at home care refers to home based nursing and rehabilitation services designed to prevent hospital admissions or facilitate early discharge. Shepperd et al randomised patients recovering from selected surgical and medical conditions to home care or ordinary inpatient care to investigate whether length of stay could be reduced and costs saved. There were few differences in outcome measures and no differences in total health care costs between the two groups [223,224]. Richards et al used a similar design for early discharge of stable elderly medical patients. Again, there was no difference in the outcome measures, but over 3 months the mean total health care costs for home patients was £2,516 and for inpatients £3,292. Because these estimates were made from incomplete datasets, statistical tests could not be performed. However, a sensitivity analysis seems to indicate that home care was less costly even if the cost of inpatient care had been over-estimated [225,226]. The opposite conclusions of these two studies indicate that substitution of lower level care for hospital care does not guarantee cost reductions. Although it may be possible to obtain savings, this probably depends crucially on how the substitution is organised and on which patient groups are targeted.

6.6 Final remarks

Our finding that 24% of the admissions were inappropriate is similar to the results of other studies. At present, general practitioners have the main responsibility for deciding who will be admitted. Factors other than considerations of the patient's health may influence these decisions. The practice of so-called "defensive medicine" means that clinicians seek to defend themselves against accusations of malpractice by being overtly cautious. However, as discussed in section 6.3.2.6, the general practitioners' specificity for predicting appropriateness is probably good, at least for emergency admissions. Since little is known about the sensitivity of these predictions, the important issue may be how many patients suffer health loss because of not being admitted when they should have been.

Contrary to common belief, the results of this study suggest that little is gained by increasing the efforts to detect inappropriate admissions at the start of the hospital stay. We suspected that inappropriate admissions could be identified by a better consideration of information available at admission than is possible for relatively inexperienced admitting physicians. To investigate whether this was possible in principle, we recruited board-certified specialists with long experience, provided them with all available information about the patients, and, in addition, the opportunity to discuss difficult cases. However, the clinical information available at the time of admission was not sufficient for making good predictions of whether a patient would benefit from his hospital stay. Presumably, admissions which could easily have been identified as inappropriate had already been sorted out by the traditional gatekeepers.

As far as we know, no study has previously tried to predict inappropriate admissions or to estimate their costs. Assessments of rates of inappropriate admissions have generally ignored the fact that they must be identified before resources are spent to obtain cost reductions. The results of this study emphasise the important role of clinical uncertainty as a determinant of cost. Clinicians keep the threshold for admission low to ensure that most of those who will benefit, are admitted. The potential savings obtained by raising this threshold were small compared to the health losses. In addition to the poor predictions, this was caused by a lower mean cost for the inappropriate admissions because of a shorter mean length of stay. This suggests that, even in today's system, these patients are identified and discharged

after an initial diagnostic work-up. Perfect identification of inappropriate admissions would not have saved more than 12% of the total costs.

Accordingly, in the investigated department, we were not able to demonstrate that savings could have been obtained by trying to reduce inappropriate admissions, and it can be discussed whether it is correct to label these admissions "inappropriate" at all. An important question is to what extent this result can be generalised to other departments of internal medicine. In a study of the rate of emergency admissions to such departments in Norway, the investigated department was found to have a lower rate than others [210]. One could speculate that this implies that it was more difficult to obtain savings in our hospital than in the other hospitals in this study, none of which were teaching hospitals. Non-teaching hospitals usually have a higher percentage of emergency admissions than university hospitals. In the present study, the percentage of inappropriate emergency admissions was lower than elective (19 vs. 34). In addition, it was more difficult for the panels to identify these than the inappropriate elective admissions. This suggests that it would also have been difficult to achieve cost reductions in non-teaching hospitals with higher percentages of emergency admissions. Higher cost savings relative to health losses could also have resulted from higher costs of inappropriate admissions relative to the appropriate. This was considered in the sensitivity analysis of Paper 5, but the savings remained modest even under this assumption. Consequently, it is questionable whether other departments of internal medicine would have found it more worthwhile to reduce the number admissions on the basis of predictions of appropriateness.

7. Conclusions

- *Primary conclusions*

In the investigated department of internal medicine, clinical judgement was unsuccessful in identifying inappropriately admitted patients at the time of admission. The most important reason for this was probably uncertainty about diagnosis and the effect of planned treatment.

Costs could have been saved by excluding admissions predicted to be inappropriate. However, this would have resulted in loss of a high percentage of the total health benefits. When compared to other interventions considered to be cost-effective, these losses were high relative to the savings.

- *Secondary conclusions*

As judged by the expert panels, the health benefits were unevenly distributed across the patients. A few patients had high gains corresponding to life-saving treatment, whereas the majority had low or no benefit. Diagnosis was the most important determinant of health benefit. Age had little effect. About one quarter of the admissions were classified as inappropriate. The mean cost of the inappropriate admissions was less than half that of the appropriate, and they represented only 12% or the total costs.

When assessing health-related benefits, the agreement between the two expert panels was fair to good. The agreement about non-health-related benefits was poor.

8. Policy implications

8.1 For departments of internal medicine to which our results can be generalised

- Based on our present knowledge, caution should be observed when attempting to reduce admissions by using clinical judgement for predicting inappropriate admissions. While it may be possible to obtain modest cost reductions, these will probably be low compared to the health losses.

8.2 For other departments and other sectors of health care

- Strategies for saving resources by limiting access to care according to expected health benefit should not be implemented without assessing their actual effects on both health benefits and costs.

9. Suggestions for further research

9.1 Evaluation of primary care physicians' decisions to admit patients to hospital

The quality of the expert panels' predictions of appropriateness in the present study was poor. As discussed in section 6.3.2.6, little is known about the ability of primary care physicians to identify patients who will benefit from hospital admission. Although there is reason to believe that their specificity for predicting appropriateness may be satisfactory, the sensitivity is unknown. In other words, some of the patients seen in primary care who would have benefited from a hospital stay may not have been admitted. This is an issue that deserves closer scrutiny.

9.2 The effect of reducing length of stays in departments of internal medicine

An investigation of the effects on health benefits and costs of an intervention to reduce length of stays should be undertaken. With a clearly defined intervention, it should be possible to randomise admitted patients to the intervention or a control group. Different kinds of interventions are possible, e.g. an effort to co-ordinate the service of other departments for patients staying in the department of internal medicine. It is well known that much time is lost while waiting for the response to referrals to other departments.

9.3 The effect of better planning of elective admissions

The effects on health benefit and costs of better planning of elective admissions to the department of internal medicine should be investigated. An intervention should be made to obtain more information about the patients before admission to allow better predictions of health benefit. If this could be achieved, our results indicate a potential for cost reductions. The costs incurred by the intervention in other sectors of health care would have to be estimated.

9.4 Investigation of admissions to a department of surgery

While our results may probably be generalised to other departments of internal medicine, at least in Norway, it is an open question whether different results would have been obtained in other types of departments. Because departments of surgery

also account for a high percentage of all hospital admissions, it would be of great value to carry out a similar study in this setting.

References

1. Anonymous. Constitution of the World Health Organization. *Chronicle of WHO* 1947; **1:** 1-2.

2. OECD Health Data 1997. [CD-ROM]. OECD. OECD Publications Service.(1997)

3. Eddy DM. Clinical decision making: from theory to practice. Rationing resources while improving quality. How to get more for less . *JAMA* 1994; **272:** 817-824.

4. Brook RH. Appropriateness: the next frontier. *BMJ* 1994; **308:** 218-219.

5. Berwick DM. Eleven worthy aims for clinical leadership of health system reform. *JAMA* 1994; **272:** 797-802.

6. Dubos R. *The mirage of health.* New York: Harper and Row; 1959.

7. Illich I. *Medical Nemesis: The Expropriation of Health.* New York: Random House; 1976.

8. Carlson RJ. *The End of Medicine.* First ed. New York: John Wiley & Sons; 1975.

9. Wildavsky A. Doing better and feeling worse: The political pathology of health policy. In: Knowles JH, Ed. ***Doing better and feeling worse.*** New York: W. W. Norton & Company; 1977: 105-124.

10. Fuchs VR. Economics, health, and post-industrial society. *Milbank Mem Fund Q Health Soc* 1979; **57:** 153-182.

11. McKeown T. *The role of medicine.Dream,mirage or nemesis?* Second ed. Oxford: Basil Blackwell; 1979.

12. McKinlay JB, McKinlay SM. The questionable contribution of medical measures to the decline of mortality in the United States in the twentieth century. *Milbank Mem Fund Q Health Soc* 1977; **55:** 405-428.

13. Powles J. The effect of health services on adult male mortality in relation to the effects of social and economic factors. *Ethics Sci Med* 1978; **5:** 1-13.

14. McKinlay JB, McKinlay SM, Beaglehole R. A review of the evidence concerning the impact of medical measures on recent mortality and morbidity in the United States. *Int J Health Serv* 1989; **19:** 181-208.

15. Evans RG, Barer ML, and Marmor TR, Eds. *Why are some people healthy and others not?* New York: Walter de Gruyter; 1994.

16. Roemer MI, Schwartz JL. Doctor slowdown: effects on the population of Los Angeles County. *Soc Sci Med* 1979; **13C:** 213-218.

17. Auster R, Leveson I, Sarachek D. The production of health, an exploratory study. *Journal of Human Resources* 1969; **4:** 412-436.

18. Hadley J. *More Medical Care, Better Health? An Economic Analysis of Mortality Rates.* First ed. Washington D.C. The Urban Institute Press; 1982.

19. Levine S, Feldman JJ, Elinson J. Does Medical Care Do Any Good? In: Mechanic D, Ed. ***Handbook of Health, Health Care and the Health Professions.*** New York: Free Press; 1983: 394-404.

20. Bunker JP, Frazier HS, Mosteller F. Improving health: measuring effects of medical care. *Milbank Q* 1994; **72:** 225-258.

21. Bunker JP. Medicine matters after all. *J R Coll Physicians Lond* 1995; **29:** 105-112.

22. Beeson PB. Changes in medical therapy during the past half century. *Medicine Baltimore* 1980; **59:** 79-99.

23. Mackenbach JP, Looman CW, Kunst AE, Habbema JD, van der Maas PJ. Post-1950 mortality trends and medical care: gains in life expectancy due to declines in mortality from conditions amenable to medical intervention in The Netherlands. *Soc Sci Med* 1988; **27:** 889-894.

24. Charlton JR, Velez R. Some international comparisons of mortality amenable to medical intervention. *Br Med J Clin Res Ed* 1986; **292:** 295-301.

25. Poikolainen K, Eskola J. The effect of health services on mortality: decline in death rates from amenable and non-amenable causes in Finland, 1969-81. *Lancet* 1986; **1:** 199-202.

26. Benham L, Benham A. The impact of incremental medical services on health status, 1963-1970. In: Andersen R, Kravits J, Anderson OW, Eds. ***Equity in health services: empirical analyses in social policy.*** Cambridge, Mass. Ballinger; 1975.

27. McDermott W. Medicine: the public good and one's own. *Perspect Biol Med* 1978; **21:** 167-187.

28. Anderson OW, Morrison EM. The Worth of Medical Care: A Critical Review. *Med Care Rev* 1989; **46:** 121-155.

29. Wennberg J, Gittelsohn A. Small area variations in health care delivery. *Science* 1973; **182:** 1102-1108.

30. Wennberg JE, Freeman JL, Culp WJ. Are hospital services rationed in New Haven or over-utilised in Boston? *Lancet* 1987; **1:** 1185-1189.

31. Wennberg JE, Freeman JL, Shelton RM, Bubolz TA. Hospital use and mortality among Medicare beneficiaries in Boston and New Haven. *N Engl J Med* 1989; **321:** 1168-1173.

32. Roos NP, Roos LL. High and low surgical rates: risk factors for area residents. *Am J Public Health* 1981; **71:** 591-600.

33. Carlisle DM, Valdez RB, Shapiro MF, Brook RH. Geographic variation in rates of selected surgical procedures within Los Angeles County. *Health Serv Res* 1995; **30:** 27-42.

34. Chassin MR, Brook RH, Park RE, Keesey J, Fink A, Kosecoff J, et al. Variations in the use of medical and surgical services by the Medicare population. *N Engl J Med* 1986; **314:** 285-290.

35. Andersen TF and Mooney G, Eds. *The Challenges of Medical Practice Variations.* London: MacMillan Press; 1990.

36. Kalseth B, Solstad K, Rønningen L, Eds. *Samdata sykehus - Sammenligningsdata for somatisk fylkeshelsetjeneste 1996.* Trondheim. Norsk institutt for sykehusforskning; 1997.

37. McPherson K, Wennberg JE, Hovind OB, Clifford P. Small-area variations in the use of common surgical procedures: an international comparison of New England, England, and Norway. *N Engl J Med* 1982; **307:** 1310-1314.

38. Mjaaland M, Revhaug A, Førde OH. Perioperative use of total parenteral nutrition. Variations in guidelines and practice in three Norwegian hospitals. *Int J Technol Assess Health Care* 1995; **11:** 741-749.

39. Wennberg JE. Population illness rates do not explain population hospitalization rates. A comment on Mark Blumberg's thesis that morbidity adjusters are needed to interpret small area variations. *Med Care* 1987; **25:** 354-359.

40. Wennberg JE, Barnes BA, Zubkoff M. Professional uncertainty and the problem of supplier-induced demand. *Soc Sci Med* 1982; **16:** 811-824.

41. Wennberg JE. The paradox of appropriate care. *JAMA* 1987; **258:** 2568-2569.

42. Shwartz M, Ash AS, Anderson J, Iezzoni LI, Payne SM, Restuccia JD. Small area variations in hospitalization rates: how much you see depends on how you look. *Med Care* 1994; **32:** 189-201.

43. Park RE, Fink A, Brook RH, Chassin MR, Kahn KL, Merrick NJ, et al. Physician Ratings for Appropriate Indications for Six Medical and Surgical Procedures. *Am J Public Health* 1986; **76:** 766-772.

44. Kahn KL, Kosecoff J, Chassin MR, Flynn MF, Fink A, Pattaphongse N, et al. Measuring the Clinical Appropriateness of the Use of a Procedure. *Med Care* 1988; **26:** 415-422.

45. McClellan M, Brook RH. Appropriateness of care. A comparison of global and outcome methods to set standards. *Med Care* 1992; **30:** 565-86IS.

46. Chassin MR, Kosecoff J, Park RE, Winslow CM, Kahn KL, Merrick NJ, et al. Does inappropriate use explain geographic variations in the use of health care services? A study of three procedures. *JAMA* 1987; **258:** 2533-2537.

47. Leape LL, Park RE, Solomon DH, Chassin MR, Kosecoff J, Brook RH. Does inappropriate use explain small-area variations in the use of health care services? *JAMA* 1990; **263:** 669-672.

48. Gray D, Hampton JR, Bernstein SJ, Kosecoff J, Brook RH. Audit of coronary angiography and bypass surgery . *Lancet* 1990; **335:** 1317-1320.

49. Graboys TB, Biegelsen B, Lampert S, Blatt CM, Lown B. Results of a second-opinion trial among patients recommended for coronary angiography . *JAMA* 1992; **268:** 2537-2540.

50. Bernstein SJ, Hilborne LH, Leape LL, Fiske ME, Park RE, Kamberg CJ, et al. The appropriateness of use of coronary angiography in New York State. *JAMA* 1993; **269:** 766-769.

51. Leape LL, Hilborne LH, Park RE, Bernstein SJ, Kamberg CJ, Sherwood M, et al. The appropriateness of use of coronary artery bypass graft surgery in New York State. *JAMA* 1993; **269:** 753-760.

52. Hilborne LH, Leape LL, Bernstein SJ, Park RE, Fiske ME, Kamberg CJ, et al. The appropriateness of use of percutaneous transluminal coronary angioplasty in New York State. *JAMA* 1993; **269:** 761-765.

53. Leape LL, Hilborne LH, Schwartz JS, Bates DW, Rubin HR, Slavin P, et al. The appropriateness of coronary artery bypass graft surgery in academic medical centers. Working Group of the Appropriateness Project of the Academic Medical Center Consortium. *Ann Intern Med* 1996; **125:** 8-18.

54. Bengtson A, Herlitz J, Karlsson T, Brandrup Wognsen G, Hjalmarson A. The appropriateness of performing coronary angiography and coronary artery revascularization in a Swedish population. *JAMA* 1994; **271:** 1260-1265.

55. Kassirer JP. The quality of care and the quality of measuring it. *N Engl J Med* 1993; **329:** 1263-1265.

56. Hicks NR. Some observations on attempts to measure appropriateness of care. *BMJ* 1994; **309:** 730-733.

57. Mulley AG, Jr., Eagle KA. What is inappropriate care? *JAMA* 1988; **260:** 540-541.

58. Phelps CE. The methodologic foundations of studies of the appropriateness of medical care. *N Engl J Med* 1993; **329:** 1241-1245.

59. Gertman PM, Restuccia JD. The Appropriateness Evaluation Protocol: A Technique for Assessing Unnecessary Days of Hospital Care. *Med Care* 1981; **19:** 855-871.

60. Siu AL, Sonnenberg FA, Manning WG, Goldberg GA, Bloomfield ES, Newhouse JP, et al. Inappropriate use of hospitals in a randomized trial of health insurance plans. *N Engl J Med* 1986; **315:** 1259-1266.

61. Strumwasser I, Paranjpe NV, Ronis DL, Share D, Sell LJ. Reliability and validity of utilization review criteria. Appropriateness Evaluation Protocol, Standardized Medreview Instrument, and Intensity-Severity-Discharge criteria. *Med Care* 1990; **28:** 95-111.

62. Rishpon S, Lubacsh S, Epstein LM. Reliability of a Method of Determining the Necessity for Hospitalization Days in Israel. *Med Care* 1986; **24:** 279-282.

63. Zwarenstein MF, Benatar SR, Shaw LR. The pattern of care in the medical wards of a teaching hospital. *S Afr Med J* 1990; **77:** 460-463.

64. Santos Eggimann B. Hospital utilization reviews under field conditions: potential and improvements. *Int J Technol Assess Health Care* 1993; **9:** 514-521.

65. Bare ML, Prat A, Lledo L, Asenjo MA, Salleras L. Appropriateness of admissions and hospitalization days in an acute-care teaching hospital. *Rev Epidemiol Sante Publique* 1995; **43:** 328-336.

66. Restuccia JD, Payne SMC, Welge CH, Dayno SJ, Kreger BE, Gertman PM. *Reducing inappropriate use of inpatient medical/surgical and pediatric services.* Washington D.C. National Technical Information Service; 1986.

67. Santos Eggimann B, Paccaud F, Blanc T. Medical appropriateness of hospital utilization: an overview of the Swiss experience. *Int J Qual Health Care* 1995; **7:** 227-232.

68. Fellin G, Apolone G, Tampieri A, Bevilacqua L, Meregalli G, Minella C, et al. Appropriateness of hospital use: an overview of Italian studies. *Int J Qual Health Care* 1995; **7:** 219-225.

69. Lorenzo S, Sunol R. An overview of Spanish studies on appropriateness of hospital use. *Int J Qual Health Care* 1995; **7:** 213-218.

70. Bentes M, Gonsalves ML, Santos M, Pina E. Design and development of a utilization review program in Portugal. *Int J Qual Health Care* 1995; **7:** 201-212.

71. Payne SM. Identifying and managing inappropriate hospital utilization: a policy synthesis. *Health Serv Res* 1987; **22:** 709-769.

72. Mackintosh JM, McKeown T, Garratt FN. An examination of the need for hospital admission. *Lancet* 1961; **i:** 815-818.

73. Kurylo LL. Measuring inappropriate utilization. *Hosp Health Serv Adm* 1976; **21:** 73-89.

74. Studnicki J, Stevens CE. The impact of a cybernetic control system on inappropriate admissions. *QRB Qual Rev Bull* 1984; **10:** 304-311.

75. Restuccia JD. The evolution of hospital utilization review methods in the United States. *Int J Qual Health Care* 1995; **7:** 253-260.

76. O'Donnell J, Pilla J, Van Gemert L. Which Hospital Admissions are Appropriate? *Aust Health Rev* 1990; **12:**19-33.

77. Black N. European collaboration on appropriateness of hospital bed use: a commentary. *Int J Qual Health Care* 1995; **7:** 185-186.

78. Liberati A, Apolone G, Lang T, Lorenzo S. A European project assessing the appropriateness of hospital utilization: background, objectives and preliminary results. The Project Steering Group and the Coordinating Center. *Int J Qual Health Care* 1995; **7:** 187-199.

79. Lang T, Davido A, Logerot H, Meyer L. Appropriateness of admissions: the French experience. *Int J Qual Health Care* 1995; **7:** 233-238.

80. Victor CR, Khakoo AA. Is hospital the right place? A survey of 'inappropriate' admissions to an inner London NHS trust. *J Public Health Med* 1994; **16:** 286-290.

81. Coast J, Inglis A, Frankel S, Gray S, Peters T. Is hospital the right place? *J Public Health Med* 1995; **17:** 239-240.

82. Coast J, Peters TJ, Inglis A. Factors associated with inappropriate emergency hospital admission in the UK. *Int J Qual Health Care* 1996; **8:** 31-39.

83. Mosvold J, Nesvold H, Laake K, Mastekaasa A. Indremedisinsk akuttavdeling. *Tidsskr Nor Laegeforen* 1985; **105:** 61-65.

84. Bjørnstad E, Bergesen, Nordhaug B, Abildgaard U, Stene-Larsen G. Innleggelser i medisinsk avdeling. *Tidsskr Nor Laegeforen* 1990; **110:** 3362-3365.

85. Bergesen, Bjørnstad E, Nordhaug B, Abildgaard U, Stene-Larsen G. Indremedisin - snever eller vid spesialitet? *Tidsskr Nor Laegeforen* 1990; **110:** 3357-3358.

86. Øie BK, Fanebust R. Øyeblikkelig hjelp-innleggelser i indremedisinsk avdeling. *Tidsskr Nor Laegeforen* 1993; **113:** 836-838.

87. Sander J. Evaluering av en medisinsk avdeling ved et tredelt sykehus. *Tidsskr Nor Laegeforen* 1976; **96:** 1634-1639.

88. Seip T, Hartviksen I, Gundersen T. Medisinsk avdeling ved et 3-delt sykehus i Nord-Norge. *Tidsskr Nor Laegeforen* 1977; **97:** 470-472.

89. Syse A, Uchermann A, Førde OH. Medisinsk utbytte av sykehusopphold. *Tidsskr Nor Laegeforen* 1983; **103:** 1335-1339.

90. Siu AL, Manning WG, Benjamin B. Patient, provider and hospital characteristics associated with inappropriate hospitalization. *Am J Public Health* 1990; **80:** 1253-1256.

91. Perneger TV, Chopard P, Sarasin FP, Gaspoz JM, Lovis C, Unger PF, et al. Risk factors for a medically inappropriate admission to a Department of Internal Medicine. *Arch Intern Med* 1997; **157:** 1495-1500.

92. Restuccia J, Shwartz M, Ash A, Payne S. High hospital admission rates and inappropriate care. *Health Aff Millwood* 1996; **15:** 156-163.

93. Schroeder SA, Showstack JA, Roberts HE. Frequency and clinical description of high-cost patients in 17 acute-care hospitals. *N Engl J Med* 1979; **300:** 1306-1309.

94. Schroeder SA, Showstack JA, Schwartz J. Survival of adult high-cost patients. Report of a follow-up study from nine acute-care hospitals. *JAMA* 1981; **245:** 1446-1449.

95. Zook CJ, Moore FD. High-cost users of medical care. *N Engl J Med* 1980; **302:** 996-1002.

96. Pompei P, Charlson ME, Sax FL, Ales K, MacKenzie CR, Norton M. Trade-offs between hospital charges and patient outcomes. *Am J Med Sci* 1988; **295:** 418-423.

97. Angus DC, Sirio CA, Clermont G, Bion J. International comparisons of critical care outcome and resource consumptions. *Crit Care Clin* 1997; **13:** 389-407.

98. Detsky AS, Stricker SC, Mulley AG, Thibault GE. Prognosis, survival, and the expenditure of hospital resources for patients in an intensive-care unit. *N Engl J Med* 1981; **305:** 667-672.

99. Fedullo AJ, Swinburne AJ. Relationship of patient age to cost and survival in a medical ICU. *Crit Care Med* 1983; **11:** 155-159.

100. Oye RK, Bellamy PE. Patterns of resource consumption in medical intensive care . *Chest* 1991; **99:** 685-689.

101. Malstam J, Lind L. Therapeutic intervention scoring system (TISS)--a method for measuring workload and calculating costs in the ICU. *Acta Anaesthesiol Scand* 1992; **36:** 758-763.

102. Ruchholtz S, Nast Kolb D, Waydhas C, Stuber R, Lewan U, Schweiberer L. [Cost analysis of clinical treatment of polytrauma patients]. *Chirurg* 1995; **66:** 684-692.

103. Llodra Calvo JC, Vazquez Mata G, Bueno Cavanillas A, Delgado Rodriguez M, Galvez Vargas R. [Evaluation of the cost of an intensive medicine unit. Relationship between the cost and the severity of the disease]. *Med Clin Barc* 1994; **103:** 49-53.

104. Schapira DV, Studnicki J, Bradham DD, Wolff P, Jarrett A. Intensive care, survival, and expense of treating critically ill cancer patients. *JAMA* 1993; **269:** 783-786.

105. Holli K, Hakama M. Treatment of the terminal stages of breast cancer. *BMJ* 1989; **298:** 13-14.

106. McCall N. Utilization and costs of Medicare services by beneficiaries in their last year of life. *Med Care* 1984; **22:** 329-342.

107. Lubitz J, Prihoda R. The use and cost of Medicare in the last 2 years of life. *Health Care Financ Rev* 1984; **5:** 117-131.

108. Lubitz JD, Riley GF. Trends in Medicare payments in the last year of life. *N Engl J Med* 1993; **328:** 1092-1096.

109. Scitovsky AA. "The high cost of dying" revisited. *Milbank Q* 1994; **72:** 561-591.

110. Nord E, Hjort PF. [Health care consumption in the last year of life. Part of total health care consumption]. *Tidsskr Nor Laegeforen* 1988; **108:** 3265-3267.

111. Lithman T, Noreen D, Norlund A, Sundstrom B. [A substantial part of health resources is reserved for the dying. Cost analysis of terminal care]. *Lakartidningen* 1994; **91:** 4390-4392.

112. Riley G, Lubitz J, Prihoda R, Rabey E. The use and costs of Medicare services by cause of death. *Inquiry* 1987; **24:** 233-244.

113. Scitovsky AA. Medical care in the last twelve months of life: the relation between age, functional status, and medical care expenditures. *Milbank Q* 1988; **66:** 640-660.

114. Shapiro MF, Wenger NS. Rethinking utilization review. *N Engl J Med* 1995; **333:** 1353-1354.

115. Feldstein PJ, Wickizer TM, Wheeler JR. Private cost containment. The effects of utilization review programs on health care use and expenditures. *N Engl J Med* 1988; **318:** 1310-1314.

116. Scheffler RM, Sullivan SD, Ko TH. The impact of Blue Cross and Blue Shield Plan utilization management programs, 1980-1988. *Inquiry* 1991; **28:** 263-275.

117. Schwartz WB, Mendelson DN. Hospital cost containment in the 1980s. Hard lessons learned and prospects for the 1990s. *N Engl J Med* 1991; **324:** 1037-1042.

118. Wickizer TM, Wheeler JR, Feldstein PJ. Have hospital inpatient cost containment programs contributed to the growth in outpatient expenditures? Analysis of the substitution effect associated with hospital utilization review. *Med Care* 1991; **29:** 442-451.

119. Rosenberg SN, Allen DR, Handte JS, Jackson TC, Leto L, Rodstein BM, et al. Effect of utilization review in a fee-for-service health insurance plan. *N Engl J Med* 1995; **333:** 1326-1330.

120. Payne SMC, Ash A, Restuccia JD. The Role of Feedback in Reducing Medically Unnecessary Hospital Use. *Med Care* 1991; **29:**AS91-106.

121. Mozes B, Halkin H, Katz A, Schiff E, Modan B. Reduction of redundant hospital stay through controlled intervention. *Lancet* 1987; **1:** 968-969.

122. Vardi A, Modan B, Blumstein Z, Lusky A, Schiff E, Barzilay Z. A controlled intervention in reduction of redundant hospital days. *Int J Epidemiol* 1996; **25:** 604-608.

123. Teno J, Lynn J, Connors AF, Jr., Wenger N, Phillips RS, Alzola C, et al. The illusion of end-of-life resource savings with advance directives. SUPPORT Investigators. Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatment. *J Am Geriatr Soc* 1997; **45:** 513-518.

124. Teno JM, Murphy D, Lynn J, Tosteson A, Desbiens N, Connors AF, Jr., et al. Prognosis-based futility guidelines: does anyone win? SUPPORT Investigators. Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatment. *J Am Geriatr Soc* 1994; **42:** 1202-1207.

125. The SUPPORT Investigators. A controlled trial to improve care for seriously ill hospitalized patients. The study to understand prognoses and preferences for outcomes and risks of treatments (SUPPORT). The SUPPORT Principal Investigators. *JAMA* 1995; **274:** 1591-1598.

126. Emanuel EJ, Emanuel LL. The economics of dying. The illusion of cost savings at the end of life. *N Engl J Med* 1994; **330:** 540-544.

127. Torrance GW. Measurement of health state utilities for economic appraisal. *J Health Economics* 1986; **5:** 1-30.

128. Torrance GW. Social preferences for health states: An empirical evaluation of three measurement techniques. *Socio-Econ Plan Sci* 1976; **10:** 128-136.

129. Nord E. *Bedømming av pasienters livskvalitet.* Oslo. Statens Institutt for Folkehelse; 1992.

130. Torrance GW, Feeny D. Utilities and quality-adjusted life years. *Int J Technol Assess Health Care* 1989; **5:** 559-575.

131. Smith A. Qualms about QALYs. *Lancet* 1987; **1:** 1134-1136.

132. Harris J. QALYfying the value of life. *J Med Ethics* 1987; **13:** 117-123.

133. La Puma J, Lawlor EF. Quality-adjusted life-years. Ethical implications for physicians and policymakers. *JAMA* 1990; **263:** 2917-2921.

134. Burrows C, Brown K. QALYs for resource allocation: probably not and certainly not now. *Aust J Public Health* 1993; **17:** 278-286.

135. Nord E. The QALY - A measure of social value rather than individual utility? *Health Econ* 1994; **3:** 89-93.

136. Nord E, Richardson J, Street A, Kuhse H, Singer P. Maximizing health benefits vs egalitarianism: an Australian survey of health issues. *Soc Sci Med* 1995; **41:** 1429-1437.

137. Mehrez A, Gafni A. Quality-adjusted life years, utility theory, and healthy-years equivalents. *Med Decis Making* 1989; **9:** 142-149.

138. Mehrez A, Gafni A. The Healthy-years Equivalents: How to Measure Them Using the Standard Gamble Approach. *Med Decis Making* 1991; **11:** 140-146.

139. Johannesson M, Pliskin JS, Weinstein MC. Are healthy-years equivalents an improvement over quality-adjusted life years?. *Med Decis Making* 1993; **13:** 281-286.

140. Johannesson M. QALYs, HYEs and individual preferences--a graphical illustration. *Soc Sci Med* 1994; **39:** 1623-1632.

141. Gafni A, Birch S. Preferences for outcomes in economic evaluation: an economic approach to addressing economic problems. *Soc Sci Med* 1995; **40:** 767-776.

142. Johannesson M. The ranking properties of healthy-years equivalents and quality-adjusted life-years under certainty and uncertainty. *Int J Technol Assess Health Care* 1995; **11:** 40-48.

143. Churchill DN, Morgan J, Torrance GW. Quality of life in end-stage renal disease. *Peritoneal Dialysis Bulletin* 1984; **4:** 20-23.

144. Churchill DN, Torrance GW, Taylor DW, Barnes CC, Ludwin D, Shimizu A, et al. Measurement of quality of life in end-stage renal disease: the time trade-off approach. *Clin Invest Med* 1987; **10:** 14-20.

145. Churchill DN, Wallace JE, Ludwin D, Beecroft ML, Taylor DW. A comparison of evaluative indices of quality of life and cognitive function in hemodialysis patients. *Control Clin Trials* 1991; **12:** 159S-167S.

146. Anonymous. Association between recombinant human erythropoietin and quality of life and exercise capacity of patients receiving haemodialysis. Canadian Erythropoietin Study Group. *BMJ* 1990; **300:** 573-578.

147. Russell JD, Beecroft ML, Ludwin D, Churchill DN. The quality of life in renal transplantation--a prospective study. *Transplantation* 1992; **54:** 656-660.

148. Berwick DM, Weinstein MC. What do patients value? Willingness to pay for ultrasound in normal pregnancy. *Med Care* 1985; **23:** 881-893.

149. Fink A, Kosecoff J, Chassin M, Brook RH. Consensus Methods: Characteristics and Guidelines for Use. *Am J Public Health* 1984; **2:** 979-983.

150. Jones J, Hunter D. Consensus methods for medical and health services research. *BMJ* 1995; **311:** 376-380.

151. Cantrill JA, Sibbald B, Buetow S. The Delphi and nominal group techniques in health services research. *Int J Pharm Pract* 1996; **4:** 67-74.

152. Hotvedt R, Kristiansen IS, Førde OH, Thoner J, Almdahl SM, Bjørsvik G, et al. Which groups of patients benefit from helicopter evacuation? *Lancet* 1996; **347:** 1362-1366.

153. Drummond MF, O'Brien B, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes.* Second ed. Oxford: Oxford Medical Publications; 1997.

154. Anonymous. *Regionsykehuset i Tromsø - Årsberetning 1993.* Tromsø. Regionsykehuset i Tromsø; 1994.

155. Slåttebrekk OV. *Hva koster pasienten?* Trondheim. Norsk institutt for sykehusforskning; 1990.

156. Efron B. *The Jackknife, the Bootstrap and Other Resampling Plans.* Philadelphia: Society for Industrial and Applied Mathematics; 1982.

157. Efron B, Tibshirani RJ. *An Introduction to the Bootstrap.* New York: Chapman & Hall; 1993.

158. Fleiss JL. *Statistical Methods of Rates and Proportions.* 2nd ed. New York: John Wiley & Sons Inc; 1981.

159. Koran LM. The reliability of clinical methods, data and judgments. *N Engl J Med* 1975; **293:** 642-646.

160. Goldman RL. The reliability of peer assessments of quality of care. *JAMA* 1992; **267:** 958-960.

161. Mather HG, Pearson NG, Read KL, Shaw DB, Steed GR, Thorne MG, et al. Acute myocardial infarction: home and hospital treatment. *Br Med J* 1971; **3:** 334-338.

162. Hill JD, Hampton JR, Mitchell JR. A randomised trial of home-versus-hospital management for patients with suspected myocardial infarction. *Lancet* 1978; **1:** 837-841.

163. Metcalfe MA, Sloggett A, McPherson K. Mortality among appropriately referred patients refused admission to intensive-care units. *Lancet* 1997; **350:** 7-12.

164. Ellis J, Mulligan I, Rowe J, Sackett DL. Inpatient general medicine is evidence based. *Lancet* 1995; **346:** 407-410.

165. Michaud G, McGowan JL, van der Jagt R, Wells G, Tugwell P. Are therapeutic decisions supported by evidence from health care research? *Arch Intern Med* 1998; **158:** 1665-1668.

166. Wennberg DE, Lucas FL, Birkmeyer JD, Bredenberg CE, Fisher ES. Variation in carotid endarterectomy mortality in the Medicare population: trial hospitals, volume, and patient characteristics . *JAMA* 1998; **279:** 1278-1281.

167. Kahneman D, Slovic P, and Tversky A, Eds. *Judgment under uncertainty: Heuristics and biases.* Cambridge: Cambridge University Press; 1982.

168. Brannen AL, Godfrey LJ, Goetter WE. Prediction of outcome from critical illness. A comparison of clinical judgment with a prediction rule. *Arch Intern Med* 1989; **149:** 1083-1086.

169. Kruse JA, Thill Baharozian MC, Carlson RW. Comparison of clinical assessment with APACHE II for predicting mortality risk in patients admitted to a medical intensive care unit. *JAMA* 1988; **260:** 1739-1742.

170. Perkins HS, Jonsen AR, Epstein WV. Providers as predictors: using outcome predictions in intensive care. *Crit Care Med* 1986; **14:** 105-110.

171. McClish DK, Powell SH. How well can physicians estimate mortality in a medical intensive care unit? *Med Decis Making* 1989; **9:** 125-132.

172. Poses RM, Bekes C, Copare FJ, Scott WE. The answer to "What are my chances, doctor?" depends on whom is asked: prognostic disagreement and inaccuracy for critically ill patients. *Crit Care Med* 1989; **17:** 827-833.

173. Lee KL, Pryor DB, Harrell FE, Jr., Califf RM, Behar VS, Floyd WL, et al. Predicting outcome in coronary disease. Statistical models versus expert clinicians. *Am J Med* 1986; **80:** 553-560.

174. Kong DF, Lee KL, Harrell FE, Jr., Boswick JM, Mark DB, Hlatky MA, et al. Clinical experience and predicting survival in coronary disease. *Arch Intern Med* 1989; **149:** 1177-1181.

175. Biörck G, Erhardt LR, Lindberg G. Prediction of Survival in Patients with Acute Myocardial Infarction. *Acta Med Scand* 1979; **206:** 165-168.

176. Gilpin EA, Olshen RA, Chatterjee K, Kjekshus J, Moss AJ, Henning H, et al. Predicting 1-year outcome following acute myocardial infarction: physicians versus computers. *Comput Biomed Res* 1990; **23:** 46-63.

177. Poses RM, Smith WR, McClish DK, Huber EC, Clemo FL, Schmitt BP, et al. Physicians' survival predictions for patients with acute congestive heart failure. *Arch Intern Med* 1997; **157:** 1001-1007.

178. Mackillop WJ, Quirt CF. Measuring the accuracy of prognostic judgments in oncology. *J Clin Epidemiol* 1997; **50:** 21-29.

179. Heyse-Moore LH, Johnson-Bell VE. Can doctors accurately predict the life expectancy of patients with terminal cancer? *Palliative Medicine* 1987; **1:** 165-166.

180. Parkes CM. Accuracy of predictions of survival in later stages of cancer. *Br Med J* 1972; 1972; **2:** 29-31.

181. Evans C, McCarthy M. Prognostic uncertainty in terminal care: can the Karnofsky index help? *Lancet* 1985; **1:** 1204-1206.

182. Forster LE, Lynn J. Predicting life span for applicants to inpatient hospice. *Arch Intern Med* 1988; **148:** 2540-2543.

183. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ* 1996; **312:** 71-72.

184. Wright JC, Weinstein MC. Gains in life expectancy from medical interventions--standardizing data on outcomes . *N Engl J Med* 1998; **339:** 380-386.

185. Classen DC, Pestotnik SL, Evans RS, Lloyd JF, Burke JP. Adverse drug events in hospitalized patients. Excess length of stay, extra costs, and attributable mortality. *JAMA* 1997; **277:** 301-306.

186. Bates DW, Spell N, Cullen DJ, Burdick E, Laird N, Petersen LA, et al. The costs of adverse drug events in hospitalized patients. Adverse Drug Events Prevention Study Group. *JAMA* 1997; **277:** 307-311.

187. Brennan TA, Leape LL, Laird NM, Hebert L, Localio AR, Lawthers AG, et al. Incidence of adverse events and negligence in hospitalized patients. *N Engl J Med* 1991; **324:** 370-376.

188. Lakshmanan MC, Hershey CO, Breslau D. Hospital admissions caused by iatrogenic disease. *Arch Intern Med* 1986; **146:** 1931-1934.

189. Steel K, Gertman PM, Crescenzi C, Anderson J. Iatrogenic illness on a general medical service at a university hospital. *N Engl J Med* 1981; **304:** 638-642.

190. Lefevre F, Feinglass J, Potts S, Soglin L, Yarnold P, Martin GJ, et al. Iatrogenic complications in high-risk, elderly patients. *Arch Intern Med* 1992; **152:** 2074-2080.

191. Ebbesen J, Buajordet I, Erikssen J, Svaar H, Brors O, Hilberg T. [Drugs as a cause of death. A prospective quality assurance project in a department of medicine]. *Tidsskr Nor Laegeforen* 1995; **115:** 2369-2372.

192. Bjerkhoel FJ, Sander J, Sorensen R. [The use of drugs as the cause of admission to a department of internal medicine]. *Tidsskr Nor Laegeforen* 1986; **106:** 587-590.

193. Laake K, Kramer S, Nordal A, Borchgrevink CF. [Side effects of drugs leading to hospitalization in emergency medical departments. How good is the reporting system?]. *Tidsskr Nor Laegeforen* 1985; **105:** 1792-1795.

194. Landsem IP, Sund V. *Pleievekter i DRG.* Trondheim. Norsk institutt for sykehusforskning; 1988.

195. Sackett DL, Haynes RB, Tugwell P. *Clinical Epidemiology - A Basic Science for Clinical Medicine.* Boston, Mass.: Little, Brown & Co; 1985.

196. Fessel WJ, van Brunt EE. Assessing quality of care from the medical record. *N Engl J Med* 1972; **286:** 134-138.

197. Romm FJ, Putnam SM. The Validity of the Medical Record. *Med Care* 1981; **19:** 310-315.

198. Burns RB, Moskowitz MA, Ash A, Kane RL, Finch MD, Bak SM. Self-report versus medical record functional status. *Med Care* 1992; **30:** MS85-MS95.

199. Brennan TA, Localio AR, Leape LL, Laird NM, Peterson L, Hiatt HH, et al. Identification of adverse events occurring during hospitalization. A cross-sectional study of litigation, quality assurance, and medical records at two teaching hospitals. *Ann Intern Med* 1990; **112:** 221-226.

200. O'Neil AC, Petersen LA, Cook EF, Bates DW, Lee TH, Brennan TA. Physician reporting compared with medical-record review to identify adverse medical events. *Ann Intern Med* 1993; **119:** 370-376.

201. Kosecoff J, Fink A, Brook RH, Chassin MR. The appropriateness of using a medical procedure. Is information in the medical record valid? *Med Care* 1987; **25:** 196-201.

202. Linn BS, Linn MW, Greenwald SR, Gurel L. Validity of Impairment Ratings Made from Medical Records and from Personal Knowledge. *Med Care* 1974; **12:** 363-368.

203. Ramos Cuadra A, Marion Buen J, Garcia Martin M, Fernandez Gracia J, Morata Cespedes MC, Martin Moreno L, et al. The effect of completeness of medical records on the determination of appropriateness of hospital days. *Int J Qual Health Care* 1995; **7:** 267-275.

204. Gjorup T, Hendriksen C, Bugge PM, Jensen AM. Global assessment of patients--a bedside study. II. Inter-observer variation and frequency of clinical findings. *J Intern Med* 1990; **228:** 147-150.

205. Leape LL, Park RE, Kahan JP, Brook RH. Group judgments of appropriateness: the effect of panel composition. *Qual Assur Health Care* 1992; **4:** 151-159.

206. Scott EA, Black N. Appropriateness of cholecystectomy in the United Kingdom--a consensus panel approach. *Gut* 1991; **32:** 1066-1070.

207. Coast J, Inglis A, Frankel S. Alternatives to hospital care: what are they and who should decide? *BMJ* 1996; **312:** 162-166.

208. Ayanian JZ, Landrum MB, Normand SL, Guadagnoli E, McNeil BJ. Rating the appropriateness of coronary angiography--do practicing physicians agree with an expert panel and with each other? *N Engl J Med* 1998; **338:** 1896-1904.

209. Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Engl J Med* 1978; **299:** 926-930.

210. Anonymous. *Hvis det haster..... . Faglige krav til akuttmedisinsk beredskap.* Oslo. Statens forvaltningstjeneste; 1998.

211. Nord E, Wisløff F, Hjorth M, Westin J. Cost-utility analysis of melphalan plus prednisone with or without interferon-alpha2b in newly diagnosed multiple myeloma. *Pharmacoeconomics* 1997; **12:** 89-103.

212. Kuntz KM, Tsevat J, Goldman L, Weinstein MC. Cost-effectiveness of routine coronary angiography after acute myocardial infarction. *Circulation* 1996; **94:** 957-965.

213. Hamel MB, Phillips RS, Davis RB, Desbiens N, Connors AF, Jr., Teno JM, et al. Outcomes and cost-effectiveness of initiating dialysis and continuing aggressive care in seriously ill hospitalized adults. SUPPORT Investigators. Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments. *Ann Intern Med* 1997; **127:** 195-202.

214. Tengs TO, Adams ME, Pliskin JS, Safran DG, Siegel JE, Weinstein MC, et al. Five-hundred life-saving interventions and their cost-effectiveness. *Risk Anal* 1995; **15:** 369-390.

215. Coast J, Inglis A, Morgan K, Gray S, Kammerling M, Frankel S. The hospital admissions study in England: are there alternatives to emergency hospital admission? *J Epidemiol Community Health* 1995; **49:** 194-199.

216. Taylor GJ, Song A, Moses HW, Koester DL, Mikell FL, Dove JT, et al. The primary care physician and thrombolytic therapy for acute myocardial infarction: comparison of intravenous streptokinase in community hospitals and the tertiary referral center. *J Am Board Fam Pract* 1990; **3:** 1-6.

217. Yealy DM, De Hart DA, Ellis G, Wolfson AB. A survey of observation units in the United States. *Am J Emerg Med* 1989; **7:** 576-580.

218. Aggarwal P, Wali JP, Ranganathan S, Kailash S, Kumar A, Mishra MC. Utility of an observation unit in the emergency department of a tertiary care hospital in India. *Eur J Emerg Med* 1995; **2:** 1-5.

219. Rydman RJ, Isola ML, Roberts RR, Zalenski RJ, McDermott MF, Murphy DG, et al. Emergency Department Observation Unit versus hospital inpatient care for a chronic asthmatic population: a randomized trial of health status outcome and cost. *Med Care* 1998; **36:** 599-609.

220. Graff L, Mucci D, Radford MJ. Decision to hospitalize: objective diagnosis-related group criteria versus clinical judgment. *Ann Emerg Med* 1988; **17:** 943-952.

221. Shamian J, Hagen B, Hu TW, Fogarty TE. The relationship between length of stay and required nursing care hours. *J Nurs Adm* 1994; **24:** 52-58.

222. Wachter RM, Katz P, Showstack J, Bindman AB, Goldman L. Reorganizing an academic medical service: impact on cost, quality, patient satisfaction, and education. *JAMA* 1998; **279:** 1560-1565.

223. Shepperd S, Harwood D, Jenkinson C, Gray A, Vessey M, Morgan P. Randomised controlled trial comparing hospital at home care with inpatient hospital care. I: three month follow up of health outcomes. *BMJ* 1998; **316:** 1786-1791.

224. Shepperd S, Harwood D, Gray A, Vessey M, Morgan P. Randomised controlled trial comparing hospital at home care with inpatient hospital care. II: cost minimisation analysis. *BMJ* 1998; **316:** 1791-1796.

225. Richards SH, Coast J, Gunnell DJ, Peters TJ, Pounsford J, Darlow MA. Randomised controlled trial comparing effectiveness and acceptability of an early discharge, hospital at home scheme with acute hospital care. *BMJ* 1998; **316:** 1796-1801.

226. Coast J, Richards SH, Peters TJ, Gunnell DJ, Darlow MA, Pounsford J. Hospital at home or acute hospital care? A cost minimisation analysis. *BMJ* 1998; **316:** 1802-1806.

Table 1 Appropriateness Evaluation Protocol criteria for appropriateness of admission

An admission is considered appropriate if one of the following criteria is met:

A. Severity of illness criteria
- 1 Sudden onset of unconsciousness or disorientation
- 2 Puls rate
  - A. <50 per minute
  - B. >140 per minute
- 3 Blood pressure
  - A. Systolic <90 or >200
  - B. Diastolic <60 or >120
- 4 Acute loss of sight or hearing
- 5 Acute loss of ability to move a body part
- 6 Persistent fever > 37.8 orally for more than 5 days
- 7 Active bleeding
- 8 Severe electrolyte/blood gas abnormality (any of the following):
  - A. Na<123 or >156
  - B. K<2.5 or >6.0
  - C. standard HCO3 (unless chronically abnormal) <20 or >36[1]
  - D. Blood pH <7.30 or >7.45
- 9 Acute progressive sensory, motor, circulatory or respiratory embarassment sufficient to incapacitate the patient (inability to move, feed, breathe etc.) Note: Must also meet Intensity of Service criterion simultaneously to certify. Do not use for back pain.
- 10 EKG evidence of acute ischemia; must be suspicion of a new MI.
- 11 Wound dehiscence or evisceration

B. Intensity of service
- 1 Intravenous medications and/or fluid replacement (does not include tube feedings).
- 2 Surgery or procedure scheduled within 24 hours requiring:
  - A. General or regional anesthesia
  - B. Use of equipment, facilities available only in hospital
- 3 Vital sign monitoring every 2 hours or more often (may include telemetry or bedside cardiac monitor)
- 4 Chemotherapeutic agents that require continuous observation for life threatening toxic reaction
- 5 Treatment in an ICU
- 6 Intramuscular antibiotics at least every 8 hours
- 7 Intermittent or continuous respirator use at least every 8 hours

Override options
- 8 Other services justifying appropriateness?
- 9 Criteria met, but inappropriate nevertheless?

[1] CO2 combining power used in the original criteria

From reference 66.

Table 2 Types of health benefit assessed by the expert panels

Type of health benefit	Abbreviation	Definition
Life expectancy gain	LEG	Remaining lifetime after the hospital stay minus remaining lifetime in a hypothetical situation without treatment for the patient's current health problem
Long-term gain in quality of life	LQG	Mean quality of life in the remaining lifetime after the hospital stay as measured with the time trade-off method, minus the mean quality of life in a hypothetical situation without treatment for the patient's current health problem
Gain in healthy-years equivalents	deltaHYE	Remaining healthy-years equivalents after the hospital stay minus remaining healthy-years equivalents in a hypothetical situation without treatment for the patient's current health problem
Health-related short-term quality of life gain	HSQG	Improvement in quality of life during or shortly after the hospital stay resulting from specific medical treatment or care, relative to a hypothetical situation without admission
Non-health-related short-term quality of life gain	NHSQG	Improvement in quality of life during or shortly after the hospital stay **not** resulting from specific medical treatment or care, relative to a hypothetical situation without admission

Table 3 Summary of studies of prediction of patient survival by physicians

Author [reference]	Year	Place	Subjects	N	Measure	Clinical expert	Result
Pompei et al [96]	1984	New York, USA	Admissions to medical services	549	Rating of 5 year prognosis on ordinal scale with 5 categories	Admitting resident	Mortality at 1 year: Predicted favourable prognosis 9%, unfavorable 50% ($p<0.01$)
Brannen et al [168]	1986	Birmingham, Alabama, USA	Admissions to MICU	215	Risk of in-hospital death	Critical care fellow in charge day after admission	Area under ROC-curve 0.899
Kruse et al [169]	1986	Detroit, USA	Admissions to MICU	366	Risk of in-hospital death	Physician in charge at admission	Area under ROC-curve 0.89
Detsky et al [98]	1977-79	Boston, Mass , USA	Admissions to MICU/CCU	1831	Risk of in-hospital death	Admitting house officer	Correlation cefficient between mean probability estimate and actual survival rate 0.94
Perkins et al [170]	1979-80	San Francisco, California, USA	Admissions to ICU	85	Survival at six weeks after admission	Resident or supervising ICU physician on fourth day of stay	Sensitivity for survival 87%, specificity 41%
McClish et al [171]	1983-84	Cleveland, Ohio, USA	Admissions to MICU (coronary patients excluded)	523	Risk of in-hospital death	Attending physician within 24 hours after admission	Area under ROC-curve 0.89
Poses et al [172]	1985-86	Richmond, Virginia, USA	Admissions to ICU	269	Probability of in-hospital death	Attending physician within 24 hours after admission	Area under ROC-curve 0.825 to 0.856
Lee et al [173]	1969-82	Durham, North Carolina, USA	Selected patients with significant CHD diagnosed by cardiac catheterization	350	Probability of event or survival after one and three years	Five senior faculty cardiologists making predictions from descriptive profiles	Rank correlation coefficient between predicted and observed survival at three years 0.49, between prediction and infarct-free survival 0.29
Kong et al [174]	1969-82	Durham, North Carolina, USA	Selected patients with significant CHD diagnosed by cardiac catheterization	350	Probability of event or survival after three years	49 cardiologists making predictions from descriptive profiles	Rank correlation coefficient betwen predicted and observed survival at three years 0.52, between prediction and infarct-free survival 0.32
Björck et al [175]	1975-76	Stockholm, Sweden	Consecutive patients with AMI treated in CCU	100	Length of survival after discharge from CCU	Residents and head of CCU	Subjective prediction adequately identified patients with a very poor or very good prognosis, predictions between 1 and 12 months unreliable
Gilpin et al [176]	1982	Four study centers in California and Canada	Patients with AMI	386	Probability of cardiac death by 1 year	Three experienced cardiologists, predictions made from computer-generated summaries	Mean area under ROC curve 0.785
Poses et al [177]	1990-92	Three hopsitals in the USA	Consecutive patients with acute congestive heart failure	1168	Probability of survival for 90 days and 1 year	Physician in emergency department, predicitions made on admission	Area under ROC curve 0.66 (90 day survival) and 0.63 (1 year survival)
Mackillop [178]	Published 1997	Ontario, Canada	Patients with cancer at an outpatient cancer center	98	Probability of cure, duration of survival of 39 incurable patients	Attending physician	Area under ROC curve 0.91 (probability of cure), 0.75 (3 month survival), 0.57 (1 year survival)
Heyse-Moore et al [179]	1984-85	London, UK	Consecutive patients with terminal cancer referred to hospice	50	Length of survival	Referring doctor	Correlation coefficient between predicted and actual survival 0.00; after exclusion of extreme outlier 0.23
Parkes [180]	1970-71	London, UK	Cancer patients admitted to hospice	168	Length of survival	Referring GP and hospital staff, at the time of referral, on admission and 1 week after admission	Correlation coefficient between predicted and actual survival 0.24-0.28, >=100% over- or underestimation in 53% of the cases
Evans [181]	1983-84	Bloomsbury, UK	Patients with cancer referred to terminal care support team	45	Upper and lower limit estimate of length of survival	Members of team (nurses, doctor, social worker) at time of visit	Actual survival within limits for 54%, greater than upper for 9%, less than lower for 37%. Pearsons r for first observation on each patient 0.44
Forster et al [182]	1983-85	Washington DC, USA	Consecutive cancer patients applying for hospice care	108	Length of survival	Community and university oncologists, general internist, from ten page application form	Pearson's r between predicted and actual survival 0.02 to 0.41. Average error 0.05 to 19.02 weeks

Table 4 Treatment of persons with established disease (from reference 184)

Disease and intervention	Target population	Gain in life expectancy (months)	
		Men	Women
Cardiovascular disease			
Myocardial revascularization with bypass grafting or percutaneous transluminal coronary angioplasty	Men with coronary artery disease		NA
	1 Vessel	1 - 7	
	2 Vessels	0 - 8	
	3 Vessels	4 - 14	
Routine beta-blocker therapy	55-year-old men who survive acute myocardial infarction		NA
	Low risk of recurrence	1.2	
	Medium risk of occurrence	4.1	
	High risk of recurrence	5.6	
Thrombolytic therapy with recombinant tissue plasminogen activator during suspected acute myocardial infarction	Patients with suspected acute myocardial infarction	15	
Thrombolytic therapy with recombinant tissue plasminogen activator as compared with streptokinase	Patients with suspected acute myocardial infarction		
	Inferior infarction	0.8 - 3.1	
	Anterior infarction	1.2 - 3.5	
Implantable cardioverter-defibrillator	Survivors of cardiac arrest with recurrent ventricular arrhythmias that do not respond to conventional therapy	36 - 46	
Amiodarone therapy	Survivors of cardiac arrest with recurrent ventricular arrhythmias that do not respond to conventional therapy	14 - 16	
Heart transplantation	Candidates with end-stage cardiac failure	31 - 99	
Ticlopidine as compared with aspirin	Patients at high risk for stroke	0.6	
Cancer			
Radical prostatectomy or radiation therapy, as compared with watchful waiting, with delayed hormonal therapy if needed	65-year-old men with localized prostate cancer	1 - 11	NA
Adjuvant chemotherapy	Women with breast cancer	NA	
	Node-positive		3.6
	Node-negative		7.7 - 11
Chemotherapy	Patients with extensive small-cell lung cancer	6.6 - 8.2	
Chemotherapy	Patients with advanced non-small-cell lung cancer	1.8 - 2.9	
Chemotherapy	Men with advanced testicular cancer	107	NA
Autologous bone marrow transplantation as compared with standard chemotherapy	Patients with relapsed non-Hodgkin's lymphoma	72	
Other			
Prophylaxis against Pneumocystis carinii pneumonia and toxoplasmosis	Patients with advanced HIV disease	5.3	
Prophylaxis against Mycobacterium avium complex, fungal infections, or cytomegalovirus	Patients with advanced HIV disease	0.2 - 0.3	
Elective surgery as compared with expectant management	50-year-olds with symptomatic gallstones	1.7	3.4
Interferon therapy	35-year-olds with chronic hepatitis B who are positive for hepatitis Be antigen and do not have cirrhosis	37	
Appendectomy	Patients with suspected acute appendicitis		
	Probable	9 - 31	
	Possible	2 - 5	

NA=not applicable

Table 5 Classification of possible causes of no health benefit from hospital care

- No treatment given
  - No disease present
  - Disease present
    - Disease resolves spontaneously without treatment
    - Disease does not resolve spontaneously
      - No palliative or curative treatment
      - alters course of disease favourably

- Treatment given
  - No disease present
    - Inappropriate treatment
  - Disease present
    - Inappropriate treatment
    - Appropriate treatment
      - No effect or adverse effect of treatment
        - Due to chance
        - Due to low quality of care
        - Due to poor patient compliance

Table 6 Appointments for follow-up according to appropriateness (n=422)

	Inappropriately admitted patients(%)	Appropriately admitted patients(%)	Total (%)
Patients without appointment for follow-up	78 (76)	173 (54)	251 (59)
Readmission scheduled			
Readmission for surgery	1 (1)	27 (8)	28 (7)
Readmission without surgery	4 (4)	31 (10)	35 (8)
Appointment in outpatient clinic			
Further diagnostic interventions planned	5 (5)	11 (3)	16 (4)
Ordinary follow-up	14 (14)	78 (24)	92 (22)
Total	102 (100)	320 (100)	422 (100)

$p<0.001$ for differences between the categories (chi-square=20.57, d.f.=4)

Table 7 Plans for follow-up of patients predicted to be inappropriately admitted (n=66)

	N (%)	Type of intervention
Patients without appointment for follow-up	44 (67)	
Readmission scheduled		
Readmission for surgery	2 (3)	One aortocoronary bypass, one aortic valve replacement
Readmission without surgery	3 (5)	
Appointment in outpatient clinic		
Further diagnostic interventions planned	2 (3)	One coloscopy, one myocardial scintigraphy
Ordinary follow-up	15 (23)	
Total	66 (100)	

Figure 1. Admissions included in the Tromsø Medical Department Health Benefit Study

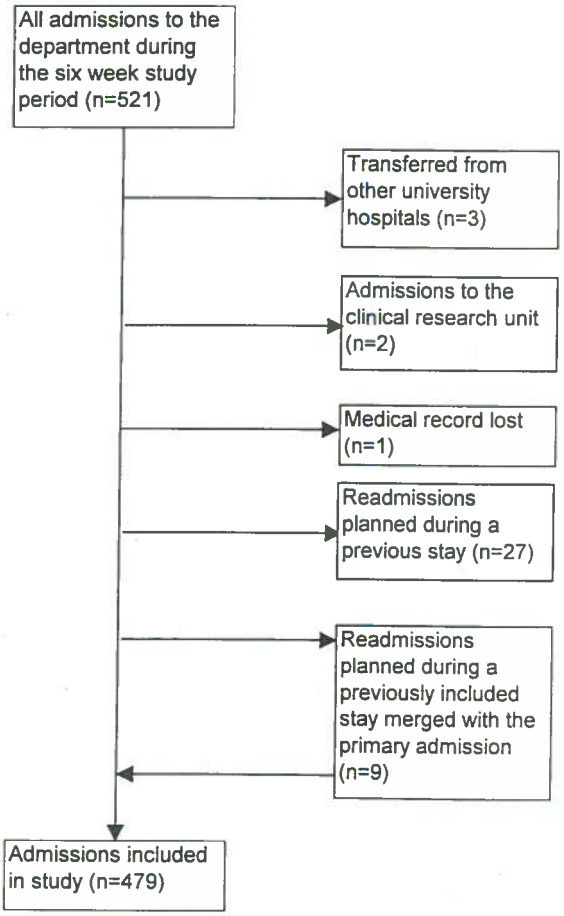


Figure 2. Design of the Tromsø Medical Department Health Benefit Study

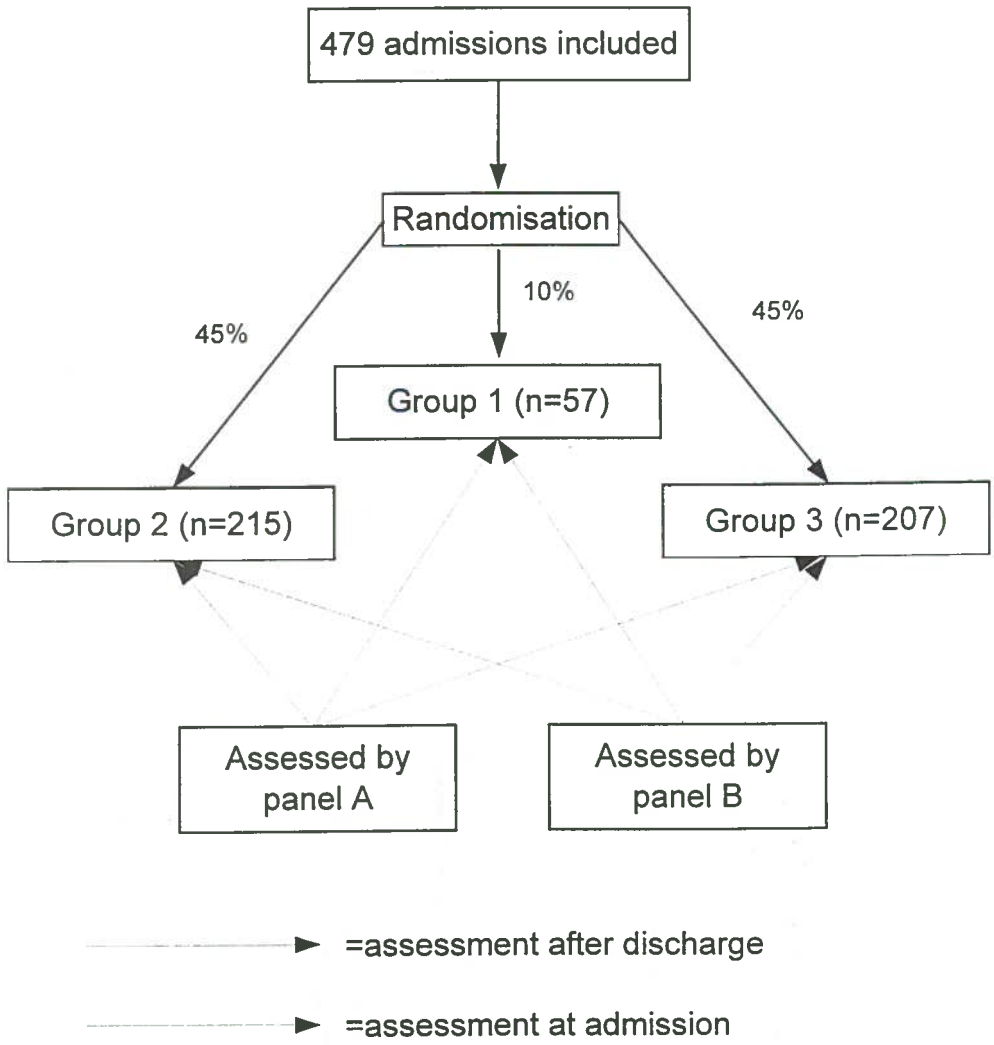
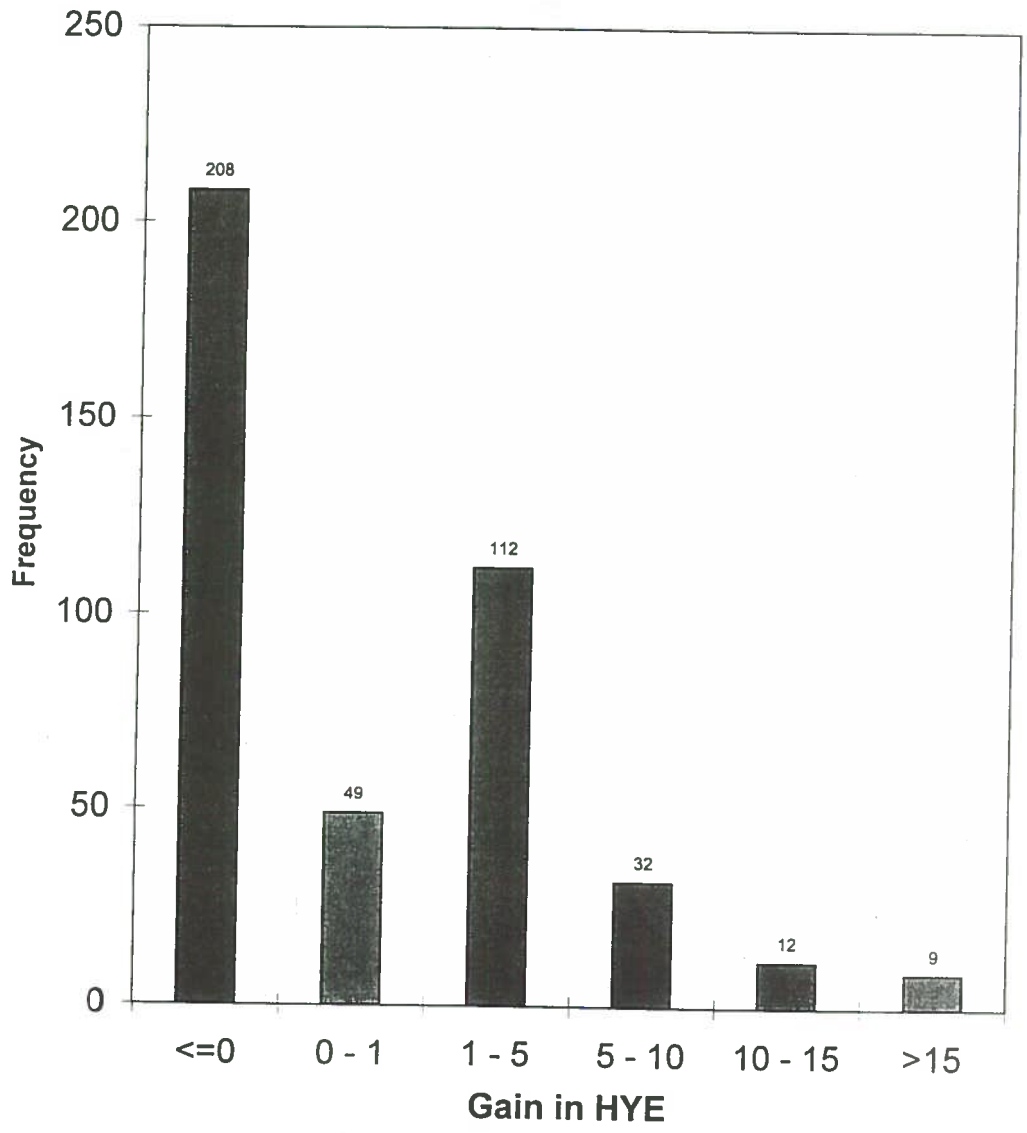
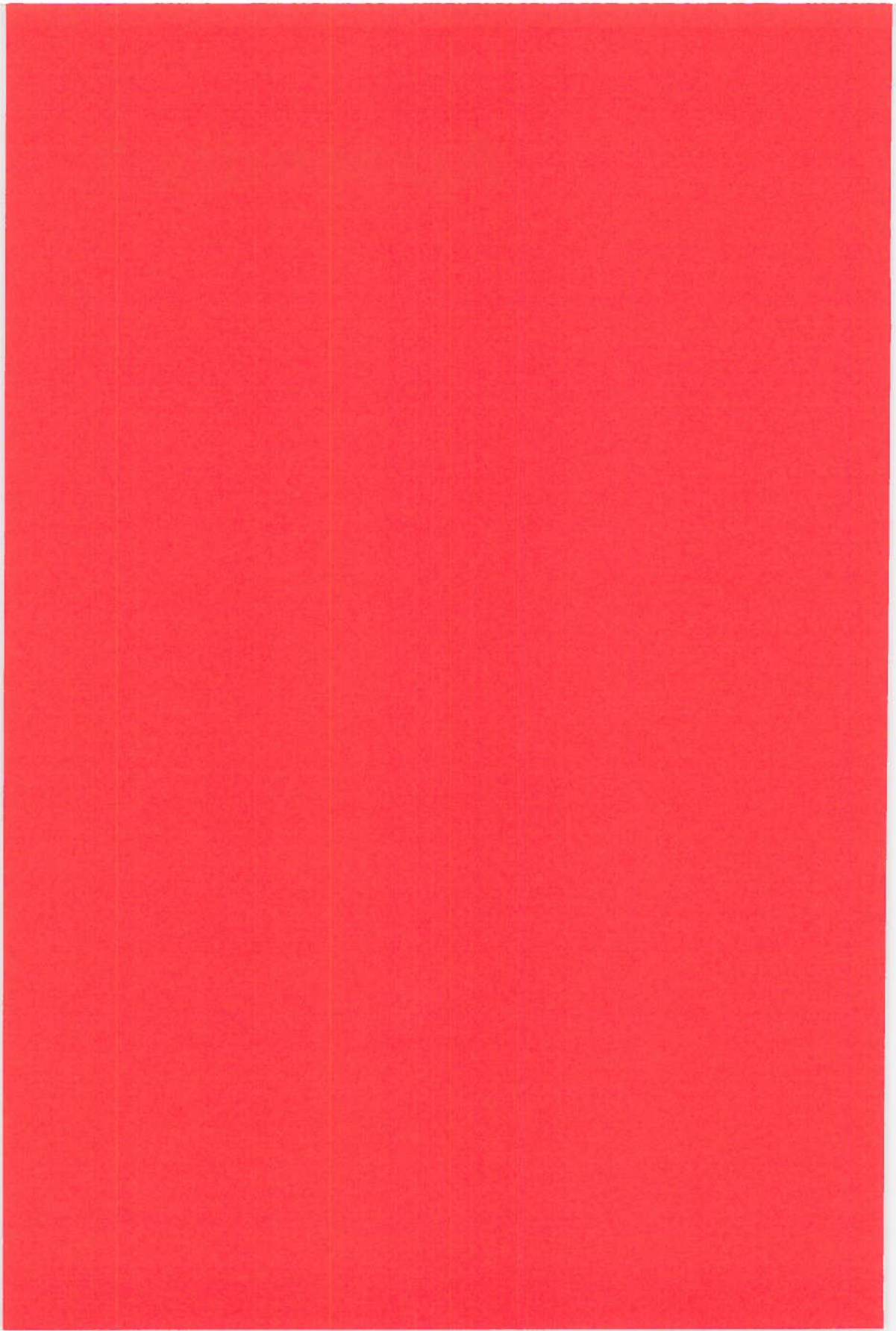


Figure 3. Distribution of gain in healthy years equivalents (HYE) from hospital admissions (n=422)



Appendix 1



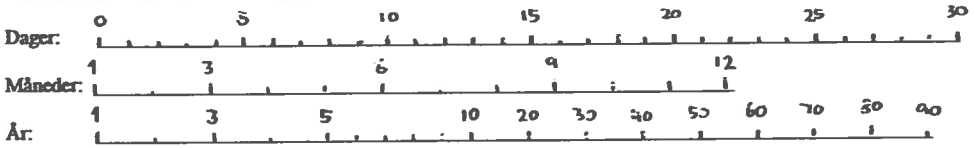
EVALUERING VED INNLEGGELSE FOR PASIENT NR. .....

Gruppe: Appendix 1

Tenk deg at pasienten ikke var blitt innlagt og ikke behandlet annet sted for det aktuelle problemet:

1. Marker med et kryss på en av skalaene nedenfor hvor lang gjenstående levetid pasienten da ville hatt i det mest sannsynlige forløpet av tilstanden. (NB! Angi gjenstående levetid, - ikke den faktisk oppnådde alder ved død!)

2. Tenk deg selv i pasientens tilstand. Bestem deg for hvor mye av den gjenstående levetiden (som du nå har markert med et kryss) du ville være villig til å ofre for å få være helt frisk fram til dødstidspunktet. Marker den gjenstående levetiden du da ville hatt med en loddrett strek på en av skalaene.

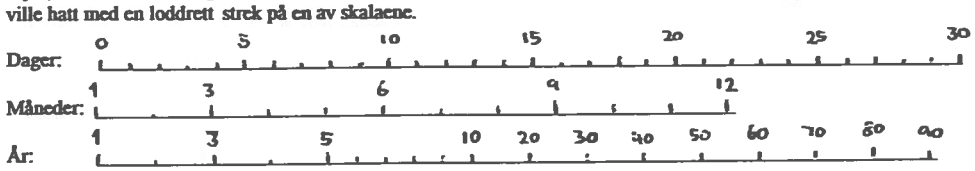


Forventet gjenstående levetid for en jevngammel frisk person er markert med en rød prikk.

Tenk deg pasientens situasjon etter dette sykehusoppholdet:

3. Marker med et kryss på en av skalaene nedenfor hvor lang gjenstående levetid pasienten vil ha oppnådd som det mest sannsynlige resultatet av oppholdet. (NB! Angi gjenstående levetid, - ikke den faktisk oppnådde alder ved død!)

4. Tenk deg selv i pasientens tilstand. Bestem deg for hvor mye av den gjenstående levetiden (som du nå har markert med et kryss) du ville være villig til å ofre for å få være helt frisk fram til dødstidspunktet. Marker den gjenstående levetiden du da ville hatt med en loddrett strek på en av skalaene.



Forventet gjenstående levetid for en jevngammel frisk person er markert med en rød prikk.

5. I hvilken grad vil dette sykehusoppholdet kunne føre til kortvarig bedring av pasientens helserelaterte livskvalitet under og like etter sykehusoppholdet i forhold til om han/hun ikke var blitt innlagt? (lindring av somatiske og psykiske plager osv.):

Sett kryss i en av rutene nedenfor.

ingen	liten	moderat	stor
☐	☐	☐	☐

6. I hvilken grad vil dette sykehusoppholdet kunne føre til kortvarig bedring av pasientens livskvalitet på annen måte under og like etter sykehusoppholdet i forhold til om han/hun ikke var blitt innlagt? (avhjelping av sosial nød, dødspleie osv.):

Sett kryss i en av rutene nedenfor.

ingen	liten	moderat	stor
☐	☐	☐	☐

7. Vil pasienten ha diagnostisk utbytte av sykehusoppholdet?:

Sett kryss i en av rutene nedenfor.

ja	nei
☐	☐

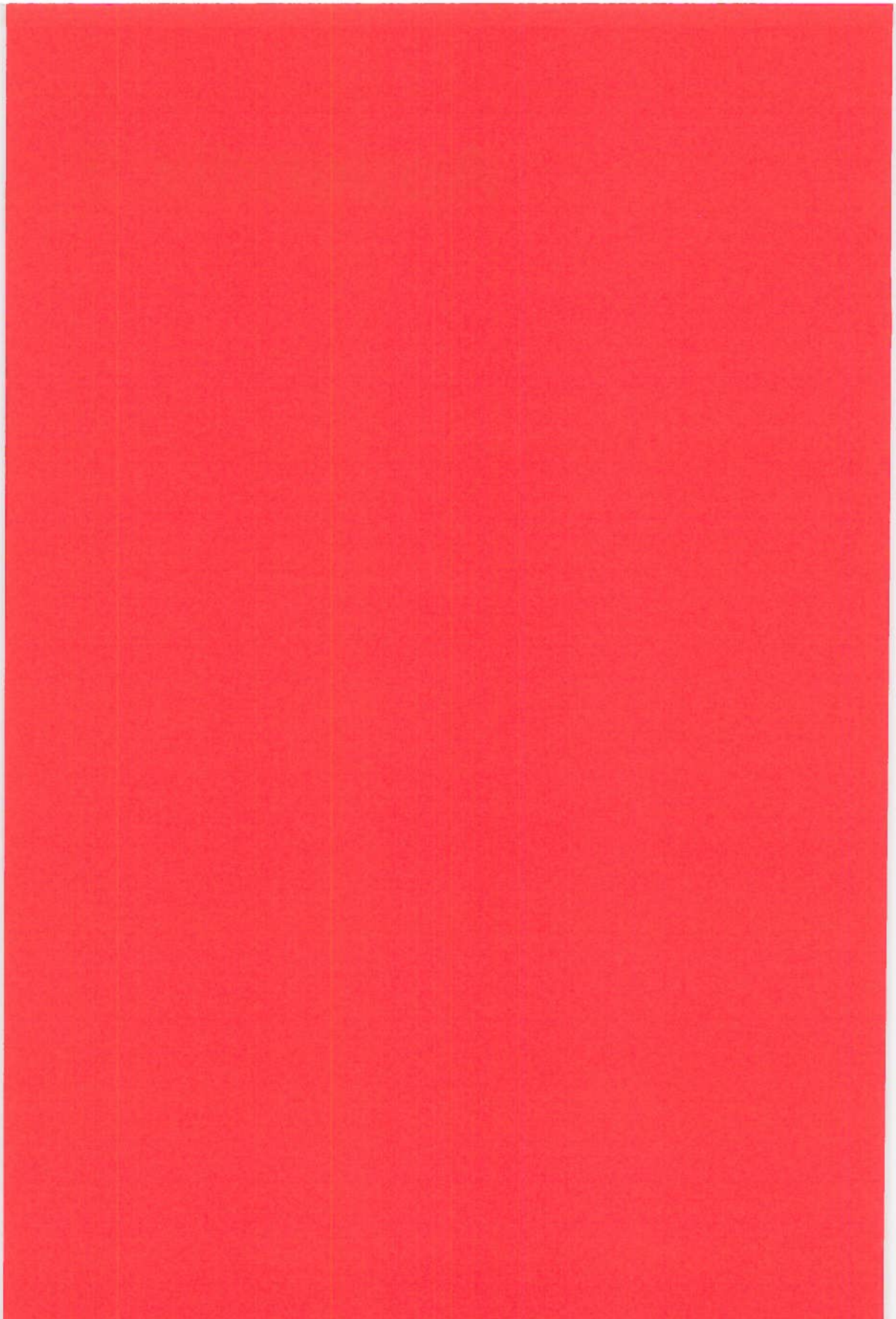
8. Forutsatt at pasienten oppnår utbytte av sykehusoppholdet, hva ville vært det laveste nødvendige omsorgsnivået for å oppnå denne gevinsten?:

Sett kryss i en av rutene nedenfor.

Primærhelsetjenesten	Medisinsk poliklinikk	Innleggelse
☐	☐	☐

PÅFØR GJERNE TEKST MED PRESISERINGER OG MARKERING AV USIKKERHET DERSOM DU ØNSKER.

Appendix 2



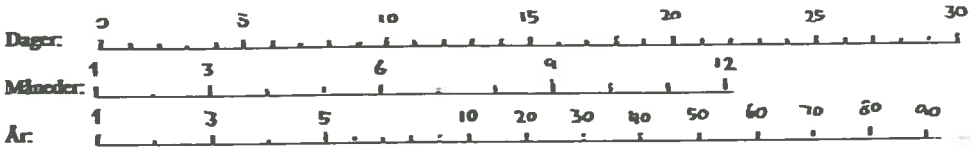
EVALUERING VED UTSKRIVELSE FOR PASIENT NR. .....

Appendix 2

Tenk deg at pasienten ikke var blitt innlagt og ikke behandlet annet sted for det aktuelle problemet:

1. Marker med et kryss på en av skalaene nedenfor hvor lang gjenstående levetid pasienten da ville hatt i det mest sannsynlige forløpet av tilstanden. (NB! Angi gjenstående levetid, - ikke den faktisk oppnådde alder ved død!)

2. Tenk deg selv i pasientens tilstand. Bestem deg for hvor mye av den gjenstående levetiden (som du nå har markert med et kryss) du ville være villig til å ofre for å få være helt frisk fram til dødstidspunktet. Marker den gjenstående levetiden du da ville hatt med en loddrett strek på en av skalaene.

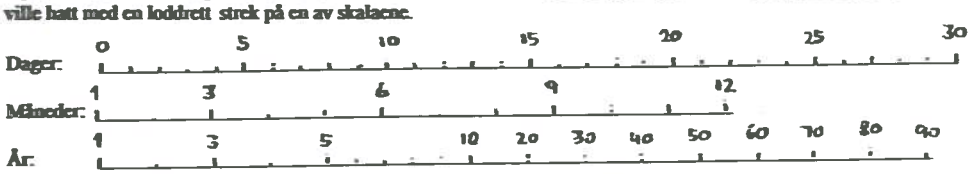


Forventet gjenstående levetid for en jevngammel frisk person er markert med en rød prikk.

Tenk deg pasientens situasjon etter dette sykehusoppholdet:

3. Marker med et kryss på en av skalaene nedenfor hvor lang gjenstående levetid pasienten har oppnådd som det mest sannsynlige resultatet av oppholdet. (NB! Angi gjenstående levetid, - ikke den faktisk oppnådde alder ved død!)

4. Tenk deg selv i pasientens tilstand. Bestem deg for hvor mye av den gjenstående levetiden (som du nå har markert med et kryss) du ville være villig til å ofre for å få være helt frisk fram til dødstidspunktet. Marker den gjenstående levetiden du da ville hatt med en loddrett strek på en av skalaene.



Forventet gjenstående levetid for en jevngammel frisk person er markert med en rød prikk.

5. I hvilken grad har dette sykehusoppholdet ført til kortvarig bedring av pasientens helserelaterte livskvalitet under og like etter sykehusoppholdet i forhold til om han/hun ikke var blitt innlagt? (lindring av somatiske og psykiske plager osv.):

Sett kryss i en av rutene nedenfor.

ingen	liten	moderat	stor
☐	☐	☐	☐

6. I hvilken grad har dette sykehusoppholdet ført til kortvarig bedring av pasientens livskvalitet på annen måte under og like etter sykehusoppholdet i forhold til om han/hun ikke var blitt innlagt? (avhjelping av sosial nød, dødspleie osv.):

Sett kryss i en av rutene nedenfor.

ingen	liten	moderat	stor
☐	☐	☐	☐

7. Har pasienten hatt diagnostisk utbytte av sykehusoppholdet?:

Sett kryss i en av rutene nedenfor.

ja	nei
☐	☐

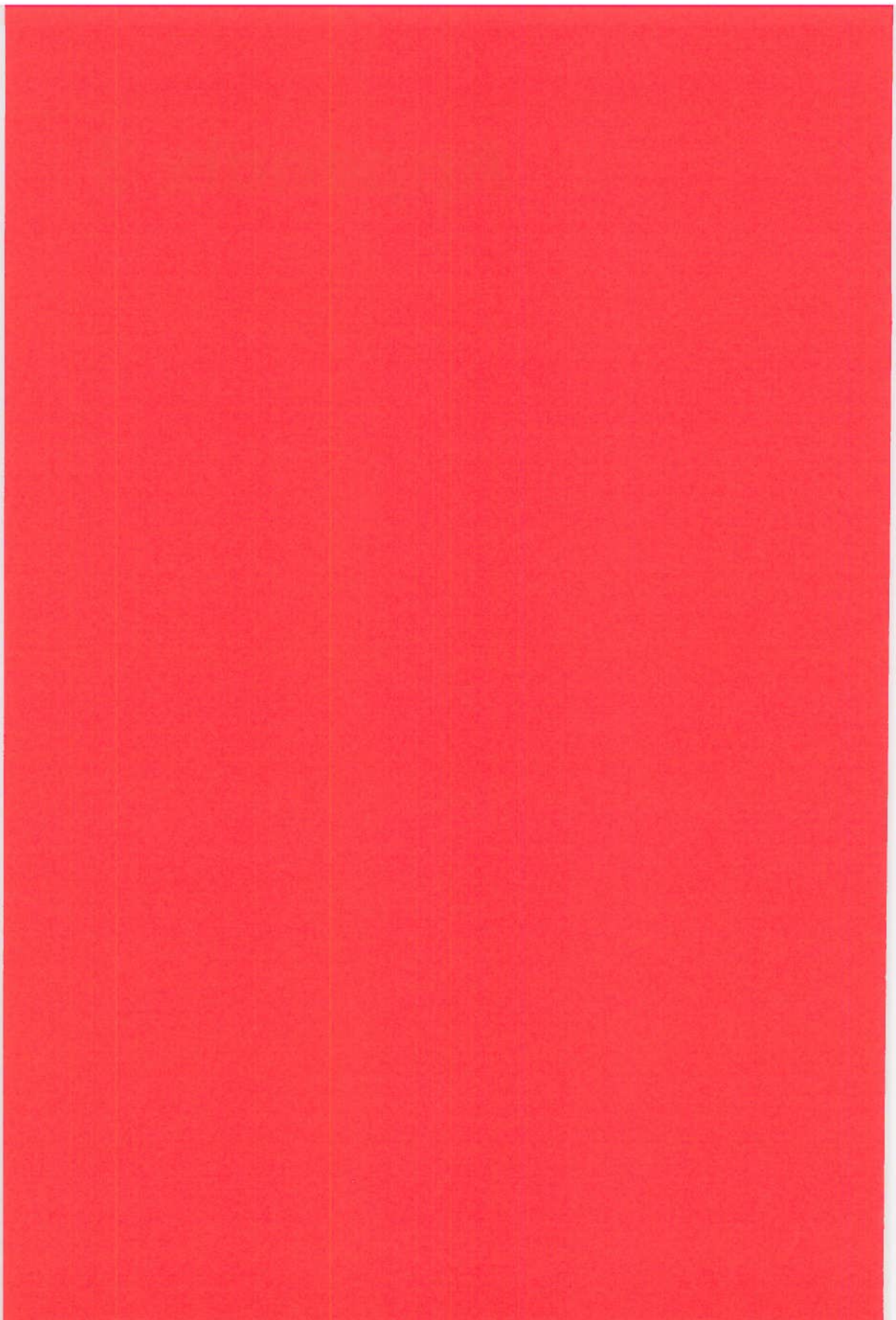
8. Forutsatt at pasienten har hatt utbytte av sykehusoppholdet, hva ville vært det laveste nødvendige omsorgsnivået for å oppnå denne gevinsten?:

Sett kryss i en av rutene nedenfor.

Primærhelsetjenesten	Medisinsk poliklinikk	Innleggelse
☐	☐	☐

PÅFØR GJERNE TEKST MED PRESISERINGER OG MARKERING AV USIKKERHET DERSOM DU ØNSKER.

Appendix 3



VEILEDNING FOR UTFYLLING AV EVALUERINGSSKJEMAENE.

I denne studien vil vi sammenligne utbyttet av et sykehusopphold med det utbyttet som forventes ved innleggelse. Hver pasient vil bli evaluert av forskjellig ekspertgruppe ved innleggelse og utskrivelse. Evalueringene i de to situasjonene foretas på samme måte: først angis forventet gjenstående levealder og livskvalitet for det tenkte tilfellet at pasienten ikke var blitt innlagt, deretter anslås på samme måte den nytten man antar pasienten vil få eller har hatt av sykehusoppholdet. I tillegg skal det både ved innleggelse og utskrivelse vurderes helsegevinst og livskvalitetsforbedring på kort sikt, diagnostisk utbytte og det laveste nødvendige omsorgsnivå for pasienten.

EVALUERING VED INNLEGGELSE FOR PASIENT NR. .....

Tenk deg at pasienten ikke var blitt innlagt og ikke behandlet annet sted for det aktuelle problemet:

Man tenker seg her at pasienten ikke var blitt innlagt og at han heller ikke hadde fått noe tilbud på lavere omsorgsnivå (vurderingen av om omsorg på lavere nivå ville vært tilstrekkelig til å gi utbytte kommer i spørsmål 8). I de fleste tilfelle vil det kunne tenkes flere muligheter for hvordan det vil gå med pasienten uten innleggelse. På bakgrunn av tilgjengelige medisinske data og sunt klinisk skjønn må du angi evalueringen i forhold til det forløpet du antar er mest sannsynlig. Det er altså ikke det verst tenkelige forløpet uten behandling det spørres etter, men det mest sannsynlige. Denne forskjellen er viktig, siden vi er vant til å begrunne innleggelser med faren for alvorlige komplikasjoner, selv om de kan være forholdsvis sjeldne.Det er heller ikke det forventede eller "gjennomsnittlige" forløpet for en gruppe av pasienter i statistisk forstand du skal fram til, men hvordan du tror det vil gå med akkurat denne pasienten! Dersom du f.eks antar at pasienten lider av tilstand x som i 50% av tilfellene fører til snarlig død, men i de resterende 50 % helbredes uten behandling, må du bestemme deg for et av disse forløpene. I en del tilfelle vil du her måtte gjette. Bli ikke frustrert av dette, fordi usikkerheten i denne situasjonen er noe av det vi vil måle med denne undersøkelsen.

1. Marker med et kryss på en av skalaene nedenfor hvor lang gjenstående levetid pasienten da ville hatt i det mest sannsynlige forløpet av tilstanden. (NB! Angi gjenstående levetid, - ikke den faktisk oppnådde alder ved død!)

Sett først et kryss på tidsskalaen for den gjenstående levetid du tror pasienten ville hatt. Dersom du tror det dreier seg om mindre enn 1 måned, setter du et kryss på den øverste skalaen. Dersom du tror det dreier seg om fra 1 måned til 1 år, setter du et kryss på den midterste skalaen. Dersom det dreier seg om mer enn et år, bruker du den nederste skalaen. Det kan settes kryss hvor som helst på skalaene; ikke bare for hele dager, måneder eller år der det er angitt markeringer. Gjennomsnittlig forventet levealder for en frisk person med samme alder og kjønn er angitt med en rød prikk som et referansepunkt. Det er selvfølgelig anledning til å anta at pasienten vil leve lenger enn gjennomsnittet, d.v.s. sette krysset til høyre for det røde punktet. Legg merke til at det er gjenstående levetid i forhold til innleggelsestidspunktet det spørres etter; ikke den alder pasienten faktisk vil oppnå.

---

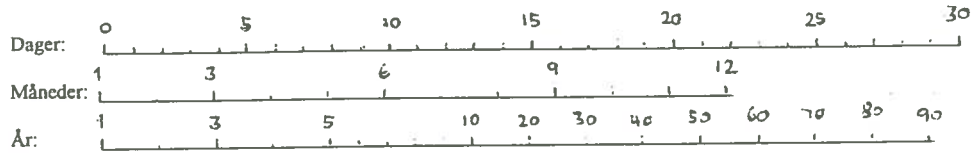
2. Tenk deg selv i pasientens tilstand. Bestem deg for hvor mye av den gjenstående levetiden (som du nå har markert med et kryss) du ville være villig til å ofre for å få være helt frisk fram til dødstidspunktet. Marker den gjenstående levetiden du da ville hatt med en loddrett strek på en av skalaene.

Når krysset som angir forventet gjenstående levetid er satt, går du videre med å angi pasienten forventede livskvalitet uten innleggelse på følgende måte: Tenk deg at pasienten fikk valget mellon to alternativer:

- ** å leve sin gjenværende levetid med den livskvalitet du antar han da vil ha, eller*
- ** å gi avkall på noe av sin gjenværende levetid i bytte mot å få være helt frisk hele tide fram til dødstidspunktet*

Jo dårligere livskvalitet pasienten antas å ha i det første tilfellet, jo kortere tid ville han/hun kunn akseptere å leve dersom han/hun kunne ha full helse. Marker med en loddrett strek de gjenstående levetid med full helse som for pasienten ville være likeverdig med å leve den faktisk gjenstående levetiden med redusert livskvalitet. Denne streken må nødvendigvis lokaliseres t venstre for eller oppå krysset som markerer gjenstående levetid. Det må taes med i beregninge at pasientens livskvalitet kan variere i løpet av den gjenstående levetiden, f.eks. at en me uhelbredelig cancer vil ha fallende livskvalitet.

Et eksempel: La oss tenke oss at en pasient med hjertesvikt har en gjenstående levealder på f. eks 2 år. Dette markeres med et kryss på den nederste skalaen. Ubehandlet vil pasienten ha my plager med dyspnoe, ødemer o.s.v. Avhengig av hvor mye dette reduserer livskvaliteten, v pasienten antagelig være villig til å gi avkall på noe av denne levetiden dersom han i stedet kunn være frisk fram til dødstidspunktet. Tenk deg i pasientens sted, og bestem deg for hvor mye a levetiden <u>du selv</u> i denne situasjonen ville være villig til å gi avkall på <u>utfra dine preferanser</u>. L oss si det dreier seg om 0,5 år (tallet er tilfeldig valgt). Du ville da oppfatte det å leve i 1,5 år so frisk, som likeverdig med å leve i 2 år med de hjertesvikt-plagene du antas å ville få. Marker 1,5 å på nederste skala med en loddrett strek. Legg merke til at dette er et tanke-eksperiment som d utfører for å få et mål for pasientens livskvalitet. Tanke-eksperimentet er <u>ikke</u> avhengig av om d faktisk eksisterer noen behandling som kan gjøre pasienten frisk bare han er villig til å of levetid.



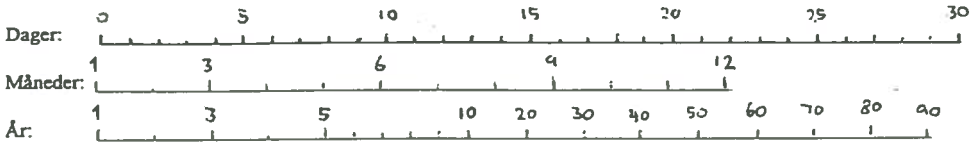
Forventet gjenstående levetid for en jevngammel frisk person er markert med en rød prikk.

Tenk deg pasientens situasjon etter dette sykehusoppholdet:

Man tenker seg det mest sannsynlige sykdomsforløpet etter utredning og behandling og gjør d samme to markeringene på tidsskalaen i denne rammen som ovenfor. Igjen må det understrekes det er det mest sannsynlige forløpet man skal fram til, i motsetning til det optimale, d.v.s. forløp med den største helsegevinsten pasienten kunne tenkes å oppnå under oppholdet. I en del tilfel vil behandling som muliggjøres av diagnostikk under oppholdet først bli gjennomført under

senere opphold (f.eks. kirurgisk behandling av nydiagnostisert neoplasme), eller strekke seg over flere senere opphold (f.eks. cytostatika-kurer for leukemi). Også i disse tilfellene vil behandlingen og behandlingsresultatet stå i et årsaksforhold til det aktuelle oppholdet, og taes med i betraktningen når man tenker seg pasientens sykdomsforløp. Et eksempel: Pasienter som legges inn for utredning av coronar hjertesykdom med angiografi m.t.p. operativ behandling må vurderes i forhold til den situasjonen de vil være i etter en senere operasjon, dersom man antar at utredningen vil føre til at slik behandling er aktuell.

3. Marker med <u>et kryss</u> på en av skalaene nedenfor hvor lang gjenstående levetid pasienten vil ha oppnådd som det mest sannsynlige resultatet av oppholdet. (NB! Angi gjenstående levetid, - <u>ikke den faktisk oppnådde alder ved død!)</u>
4. Tenk deg selv i pasientens tilstand. Bestem deg for hvor mye av den gjenstående levetiden (som du nå har markert med et kryss) du ville være villig til å ofre for å få være helt frisk fram til dødstidspunktet. Marker den gjenstående levetiden du da ville hatt med en loddrett strek på en av skalaene.



Forventet gjenstående levetid for en jevngammel frisk person er markert med en rød prikk.

5. I hvilken grad vil dette sykehusoppholdet kunne føre til kortvarig bedring av pasientens helse under og like etter sykehusoppholdet i forhold til om han/hun ikke var blitt innlagt? (lindring av somatiske og psykiske plager osv.):
Sett kryss i <u>en</u> av rutene nedenfor.

ingen	liten	moderat	stor
☐	☐	☐	☐

Her er det spørsmål om kortvarig bedring av helse som følge av sykehusoppholdet, sett i forhold til om pasienten ikke var blitt innlagt. En spontan bedring av helsen uavhengig av tiltak satt i verk under oppholdet regnes derfor ikke som gevinst her. Med forbedring av helse mener vi her f.eks:

Lindring av somatiske plager(smerte, ubehag , kvalme etc., etc.)
Lindring av psykiske plager (depresjon, angst etc.)
Bedring av funksjonsevne
etc.

For å gi en pekepinn om hvilken bedring som vil svare til de fire kategoriene ovenfor, vil vi be deg kikke på EuroQol-skalaen i protokollen (vedlegg 3). En bedring av helsen er moderat dersom den omtrent tilsvarer en forbedring på et trinn på en av de 5 dimensjonene i skalaen. Bedringen betegnes som stor dersom den omtrent tilsvarer et sprang på to trinn av en av dimensjonene eller en forbedring på mer enn en av dimensjonene. Dette er bare ment som en illustrasjon på hva som legges i kategoriene ovenfor; det er ikke meningen at du skal bruke EuroQol-skalaen når dette og det neste spørsmålet besvares.

6. I hvilken grad vil dette sykehusoppholdet kunne føre til kortvarig bedring av pasientens livskvalitet på annen måte under og like etter sykehusoppholdet i forhold til om han/hun ikke var blitt innlagt? (avhjelping av sosial nød, dødspleie osv.):

Sett kryss i en av rutene nedenfor.

ingen	liten	moderat	stor
☐	☐	☐	☐

Dette spørsmålet skal omfatte alle typer kortvarig bedring av livskvaliteten som ikke omfattes av spørsmål 5. Eksempler:

Avhjelping av akutt vanskelig sosial situasjon (f.eks. akutt pleie)
Bedre forståelse av egen helsetilstand ved at man finner forklaring på symptomer, får informasjon om egen sykdom o.l.
Dødspleie
etc., etc

Spørsmålet besvares på samme måte som spørsmål 5, og også her vil vi vise til EuroQol-skalaen for å illustrere hvordan kategoriene brukes.

7. Vil pasienten ha diagnostisk utbytte av sykehusoppholdet?:

Sett kryss i en av rutene nedenfor.

ja	nei
☐	☐

Pasienten regnes å ha hatt gevinst av diagnostikk utført under oppholdet dersom resultatene av diagnostikken ville kunne gi gevinst i h.h.t spørsmål 1 - 6, eller dersom den gir forklaring på sykdomsmanifestasjoner (symptomer, tegn, patologiske prøvesvar o.l.). Avkrefting av foreslåtte diagnoser eller diagnostikk uten at det blir stilt noen diagnose, gir ikke gevinst her.

8. Forutsatt at pasienten oppnår utbytte av sykehusoppholdet, hva ville vært det laveste nødvendige omsorgsnivået for å oppnå denne gevinsten? (Spørsmålet skal besvares for alle pasienter):

Sett kryss i en av rutene nedenfor.

Primærhelsetjenesten	Medisinsk poliklinikk	Innleggelse
☐	☐	☐

Dette spørsmålet besvares for alle pasienter, uansett om man tror de vil ha utbytte eller ikke. Dersom man ikke tror pasienten vil få noe utbytte, skal men evt. krysse av for laveste kategori. Med primærhelsetjeneste menes alle typer tilbud som omfattes av helse- og sosialtjenesten i de fleste kommuner. Med medisinsk poliklinikk menes vanlig eller ø.hj.-konsultasjon på medisinsk poliklinikk. Overnatting på sykehotell regnes ikke som en del av tilbudet på medisinsk poliklinikk. Vi har alle en oppfatning av hvilke pasientkategorier som bør behandles på de forskjellige nivåene. Det er imidlertid ikke det vi her vil fram til, men hvilket nivå som ville være nødvendig i forhold til den gevinsten pasienten faktisk ventes å oppnå.

EVALUERING VED UTSKRIVELSE FOR PASIENT NR. .....

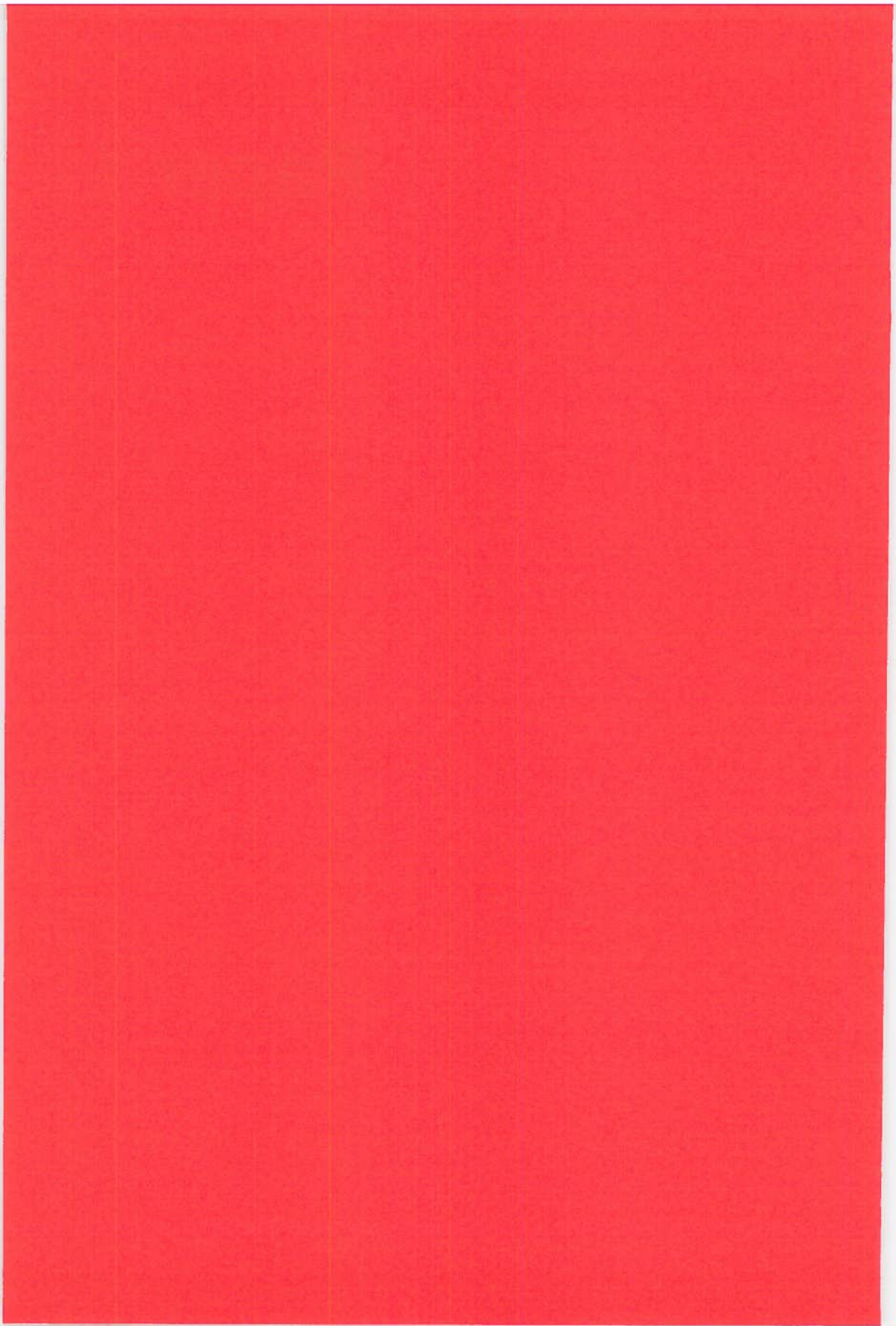
Denne evalueringen er nesten identisk med den som foretas ved innleggelse. Forskjellen er at man nå har data samlet inn under sykehusoppholdet som grunnlag for å anslå sykdomsforløpet både med og uten innleggelse. I den grad det fortsatt er flere mulige forløp, velges det mest sannsynlige. For noen pasienter vil diagnostikk under oppholdet ha muliggjort behandling som er planlagt og avtalt, men ennå ikke gjennomført under oppholdet. Effekten av slik behandling på forløpet må taes med i betraktning (f.eks operasjoner, strålebehandling, cytostatika-kurer o.a.)

PÅFØR GJERNE TEKST MED PRESISERINGER OG MARKERING AV USIKKERHET DERSOM DU ØNSKER, *men husk på at alle spørsmålene på skjemaene skal besvares.*

DIVERSE

Evalueringene må returneres til prosjektkoordinator fortløpende ettersom de fylles ut. Både evalueringsskjemaet og pasientopplysningene skal returneres. Dersom det oppstår praktiske problemer m.h.t. evalueringen, kan prosjektkoordinator kontaktes (kl. 0800 - 1600 tlf. 083 26000, kl. 1600 - 0800 tlf.083 xxxxx).

Paper 1



International Journal of Technology Assessment in Health Care, 12:1 (1996), 126–135.

ASSESSING HEALTH BENEFIT FROM HOSPITALIZATION

Agreement Between Expert Panels

Bjørn O. Eriksen
Sven M. Almdahl
University Hospital of Tromsø

Anne Hensrud
Kommunelegekontoret i Bardu

Steinar Jæger
Nordland Central Hospital

Ivar S. Kristiansen
University of Tromsø

Fred A. Mürer
Rana Hospital

Erik Nord
National Institute of Public Health, Norway

Jan Fr. Pape
University Hospital of Tromsø

Reidar Robertsen
Åsgård Psychiatric Hospital

Glen Thorsen
Harstad sykehus

Abstract

Agreement between two expert panels in assessing gain in life expectancy and quality of life from unselected stays in a department of internal medicine was investigated. Weighted kappa statistics of 0.45 for gain in life expectancy and 0.63 for gain in quality of life were found.

The rising cost of health care makes the optimal allocation of resources a vital issue. To find the best allocations, it is necessary to estimate the health benefit of competing

This study was supported by a grant from the Norwegian Medical Association

health care programs. Methods for the quantification of health have been developed (15), but have rarely been used for this purpose. The reason may be that the measurement of health is connected with both practical and conceptual difficulties, some of which have been discussed elsewhere (11).

Health can be defined as a function of life expectancy and some measure of the quality of life (15). Although survival and gains in life expectancy are frequently end points in clinical trials, these results are of limited value in estimating life expectancy gain from hospital care, because restrictive inclusion criteria often make it difficult to apply them to ordinary patients. Many technologies and treatment modalities have never been evaluated properly in clinical trials. Information about improvements in quality of life is even more scarce. Consequently, evaluation of health benefits from empirical data alone is not possible, and we are left with clinical judgment as the second best alternative.

Methods based on expert clinical judgment have been used in many different settings for the evaluation of health care programs. Usually, a selected group of patients is evaluated according to program-specific criteria, which makes comparisons between different programs difficult. To our knowledge, no study has so far assessed improvement in such general health measures as life expectancy and quality of life for a group of unselected patients admitted to a hospital department.

The Tromsø Medical Department Health Benefit Study relies on consensus in panels of expert clinicians to estimate gains in life expectancy and quality of life arising from hospital stays. It was designed to study health effects and resource utilization in a department of internal medicine. A major objective of the study was to identify patient groups with very low health gain. To investigate the reliability of the method, the interpanel agreement for the health measures was studied in a random sample of the patients included in the study.

MATERIAL AND METHODS

Subjects

In 1992, 4,567 patients were admitted to the Department of Internal Medicine of the University Hospital of Tromsø. During a six-week period from February 1, 1993, all admissions to this department were considered for inclusion in the Tromsø Medical Department Health Benefit Study. The groups excluded were: (a) patients transferred from other university hospitals ($n = 3$); (b) patients admitted for evaluation or continuation of treatment started during a previous stay ($n = 27$); and (c) patients admitted for inclusion in drug trials ($n = 2$). One patient was excluded because his medical record could not be found. Of the 488 remaining, nine planned readmissions were merged with the primary admission, resulting in the inclusion of a total of 479 hospital stays. Each stay was given a probability of 0.10 of being randomized to the present study for the purpose of investigating interpanel agreement. Randomization was performed with a pseudorandom number generator.

Expert Panels

Two expert panels were established. Each expert panel consisted of one internist, one surgeon, and one general practitioner. All the experts were board-certified specialists in their respective fields. None of them had any connection with the department being studied. All the hospital stays in this study were evaluated by both expert panels.

Before the study began, the experts examined and discussed the evaluation protocol thoroughly. The study was approved by the Regional Ethics Committee and

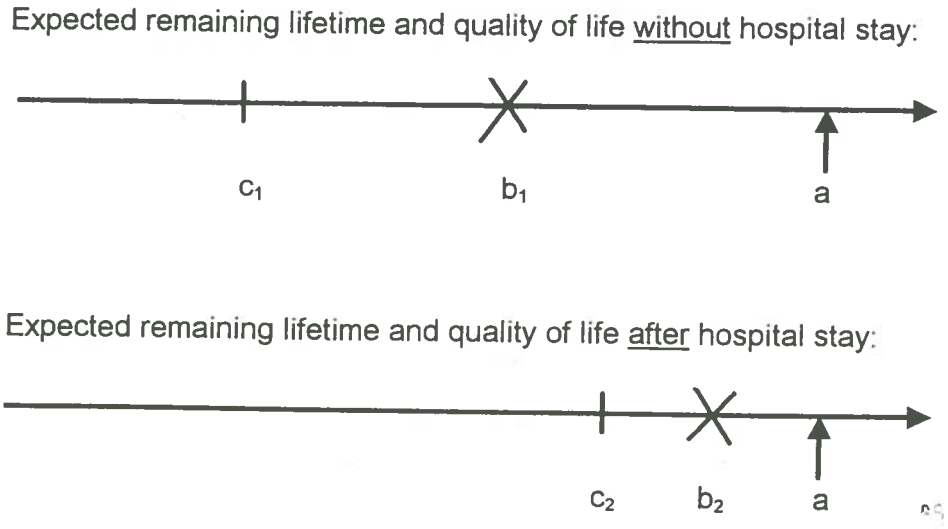


Figure 1. Time scales for assessing gain in life expectancy and quality of life. It was assumed that the patient would not have received any treatment for the current health problem had he or she not been admitted to hospital. a = Life expectancy of person of same age and sex in the general population (information given by the project coordinator); b_1 = life expectancy of the patient had he or she not been hospitalized; b_2 = life expectancy of the patient after hospital stay (life expectancy gain from hospital stay = $b_2 - b_1$); c_1 = lifetime in perfect health after having traded off time equal to $b_1 - c_1$ in the hypothetical situation without hospitalization; c_2 analogous to c_1 in the situation after this hospital stay. Mean quality of life without hospitalization = c_1/b_1. Mean quality of life after hospital stay = c_2/b_2. Gain in quality of life from hospital stay = $c_2/b_2 - c_1/b_1$.

the Norwegian Data Inspectorate. It was subsequently tested in a pilot study with 10 cases.

Data

Detailed descriptions of each hospital stay were compiled from the patients' discharge reports and medical records by the project coordinator (BOE), who is a board-certified specialist in internal medicine. The descriptions included social history, previous illnesses, current problem, medication, physical findings, results of tests, treatment during the stay, and plans for further treatment. For patients transferred to other departments in the hospital, the discharge reports from these departments were included, as was information from planned readmissions or further diagnostic procedures within 2 months of the primary admission. The summaries were blinded, both with respect to the identity of the patients and of their physicians.

Evaluation of Improvement in Life Expectancy and Quality of Life

As part of a questionnaire on the health benefit of the hospital stays, the experts assessed gain in life expectancy and quality of life. Life expectancy was recorded on two separate time scales (Figure 1): one for the patient's situation after the stay (b_2), and one for the hypothetical situation had he or she not been hospitalized or treated elsewhere (b_1). Life expectancy gain was calculated as the difference between these two assessments. The experts were given information about the life expectancy of a person of the same sex and age in the general population by a mark on the time scales (a).

Quality of life was measured with the time trade-off techniques (16). Using the same two time scales as above, the experts were asked to decide how much of their remaining lifetime they would have been willing to exchange for perfect health up to the time of death, had they been in the patient's situation. The lifetime left after this trade-off was recorded on the time scale, with the average quality of life calculated as the ratio between this quantity and the total remaining lifetime (Figure 1). A ratio of 0 corresponded to the lowest possible quality of life, i.e., coma or death; 1.0 to perfect health. This procedure was carried out for the patient's situation if he had not been hospitalized or otherwise treated (c_1), and then again for his actual situation after this hospital stay (c_2). The gain in quality of life was defined as the difference between the average quality of life in the two situations.

Consensus Criteria

The hospital stays were first evaluated by each expert individually. For each of the two expert panels, consensus between the three experts was defined to exist if: (a) the difference between the maximum and minimum life expectancy gain estimates did not exceed 25% of the average estimated life expectancy of the patient after the hospital stay; and (b) the difference between the maximum and minimum quality-of-life gain estimates did not exceed 0.20. When both criteria were met, the panel's assessment was defined as the median of the three individual assessments. Otherwise, the case was discussed in a meeting of the three members of the panel, led by the project coordinator who did not take part in the discussion. After the discussion, the experts revised their individual estimates, and the median was again taken to represent the panel's assessment, even if the consensus criteria were not met.

There was no contact between the two expert panels during the study.

Statistical Methods

To investigate the structure of agreement between the two panels, the assessments were divided into categories of low, intermediate, and high gain and tabulated against each other in a 3×3 contingency table. The log-linear model of nonhomogeneous agreement described by Tanner and Young (14) was used for finding separate parameters characterizing each category, the antilog of which we will define here as agreement parameters. The agreement parameters can be interpreted as the ratio between the modeled probability of agreement for a category and the probability expected from chance alone. A value greater than 1 indicates higher agreement than expected by chance, and a value less than 1, lower agreement. Before fitting log-linear models, sampling zeroes in the contingency tables were eliminated by calculating pseudo-Bayes estimates of the cell counts (1).

The weighted kappa statistic was calculated with the squares of the number of categories of disagreement used as weights (3). Statistical significance was set at $p < .05$.

RESULTS

Of the hospital stays included in the main part of the Tromsø Medical Department Health Benefit Study, 57 were randomized to the present investigation of interpanel agreement.

Life Expectancy

The final estimates of life expectancy gain by both expert panels are shown in Figure 2. The median difference between the assessments of panel A and B was 0.0 years

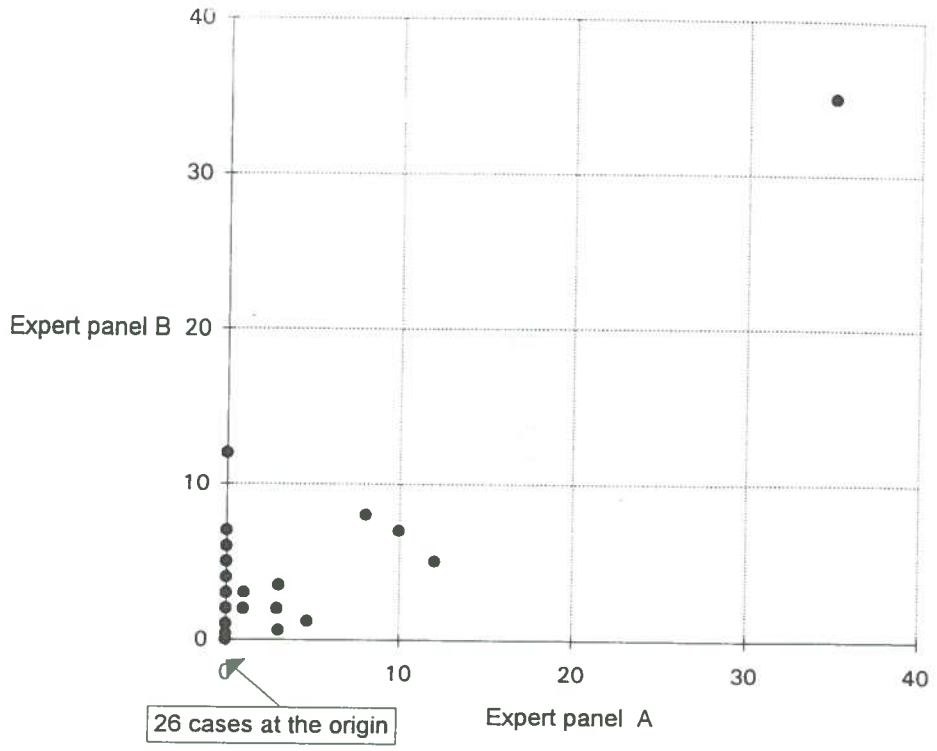


Figure 2. Life expectancy gain from hospital stays in years estimated by both expert panels (n = 57).

(range: −12.0 to 7.0 years; 5%, −5.2 years; 95%, 2.5 years). There was agreement that there would be no life expectancy gain in 26 cases (46%).

Agreement between the expert panels was analyzed with assessments categorized into low (0–0.5 years), intermediate (0.5–5 years), and high (>5 years) life expectancy gain (Table 1). The model of nonhomogeneous agreement provided an excellent fit with a log-likelihood ratio of 0.01, df = 1 (p = .91). The agreement parameter for low gain was 7.01, for intermediate gain, 3.60, and for high gain, 10.22.

The overall agreement in Table 1 was 0.67, and the weighted kappa statistic was 0.45 (95% confidence interval, 0.18–0.73).

Quality of Life

Estimates of gain in quality of life are shown in Figure 3. The median difference between the assessments of panel A and B was 0.00 (range: −0.25 to 0.35, 5%, −0.05; 95%, 0.18). There was agreement that there would be no quality of life gain in 29 cases (51%).

In the analysis of agreement, the assessments were grouped into three categories: no gain (0), intermediate gain (0–0.10), and high gain (>0.10) (Table 2). The dividing line between the middle and upper categories was set as low as 0.10, because there were very few patients with a high gain. The model of nonhomogeneous agreement fitted the data well with a log-likelihood ratio of 0.26, df = 1 (p = .61). The agreement parameter for low gain was 12.81, for intermediate gain, 0.43, and for high gain, 11.85.

Table 1. Hospital Stays Categorized According to Life Expectancy Gain by Both Expert Panels (Pseudo-Bayes Estimates of Cell Counts) (n = 57)

		Expert panel B			
		0–0.5 years	0.5–5 years	>5 years	Total
Expert panel A	0–0.5 years	29 (27.5)	15 (14.4)	3 (3.2)	47 (45.2)
	0.5–5 years	0 (0.4)	6 (6.0)	0 (0.4)	6 (6.9)
	>5 years	0 (0.4)	1 (1.4)	3 (3.2)	4 (5.0)
	Total	29 (28.3)	22 (21.8)	6 (6.9)	57 (57.0)

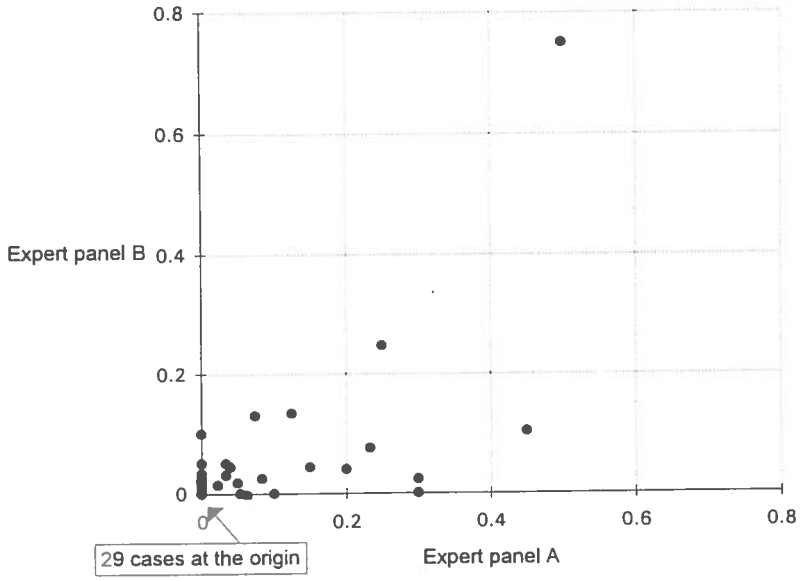


Figure 3. Quality of life gain estimated with the time trade-off method by both expert panels (n = 57).

Overall agreement in Table 2 was 0.68, and weighted kappa was 0.63 (95% confidence interval, 0.45–0.80).

DISCUSSION

The health benefit from a hospital stay is not simply the difference between health status on admission and discharge. An explicit evaluation of the patient's prognosis without hospitalization and without any other form of treatment must also be made. In this study, we assessed the quantity and quality of life in both situations to find the effect of the hospital stay, which is the difference between these two assessments.

Agreement

For both measures, the median difference between the assessments of the two expert panels was zero, and the 90% interpercentile interval quite narrow. However, inspection of the data (Figures 2 and 3) revealed poor agreement when one of the assessments

Table 2. Hospital Stays Categorized According to Quality of Life Gain as Assessed With the Time Trade-off Method by Both Expert Panels (Pseudo-Bayes Estimates of Cell Counts) (n = 57)

		Expert panel B			
		0	0–0.1	>0.1	Total
Expert panel A	0	29 (27.5)	9 (8.8)	0 (0.4)	38 (36.7)
	0–0.1	3 (3.2)	6 (6.0)	1 (1.4)	10 (10.6)
	>0.1	1 (1.4)	4 (4.2)	4 (4.2)	9 (9.7)
	Total	33 (32.1)	19 (19.0)	5 (5.9)	57 (57.0)

differed from zero. The favorable median and interpercentile intervals were a result of agreement that there would be no improvement in life expectancy for 46% and in quality of life for 51% of the patients. In addition, there was reasonable agreement on a few patients with high gain for both measures. To analyze this pattern, the assessments were divided into categories of low, intermediate, and high gain. Because the weighted kappa statistic gives no information about agreement for the separate categories of a contingency table, modeling with log-linear models was used. Models of nonhomogeneous agreement fitted the data very well, with the probability of agreement for categories of low or high gain from 7.01 to 12.81 times that expected from chance alone. From the perspective of priority setting, the ability of a method to identify patient groups with a very low or high health benefit is essential.

In a recent study, Goldman examined the interreviewer agreement of peer assessment of implicit evaluation of patient care episodes based on a review of medical records or record abstracts (7). He found only two of 12 studies with kappa values were consistently above 0.40, the conventional dividing line between agreement characterized as "poor" and "fair to good" (5). For our two measures, weighted kappa values were 0.45 and 0.63, which compare favorably.

It is also relevant to compare this type of clinical judgment with the reliability of other clinical methods. Koran (9) reviewed the interobserver agreement on clinical signs and found kappa values ranging from 0.51 for palpation of the dorsalis pedis pulse to 0.70 for interpretation of ECGs, which is only slightly better than the expert panels' judgments about health benefit. Such judgments obviously can be made with a reliability comparable with that of methods generally accepted as valuable clinical tools.

Validity

The validity of the method was not examined in this study. The gold standard would have been randomization of patients to hospital admission or no treatment followed by patient self-assessment of quality of life at regular intervals for the rest of their lives.

For ethical and practical reasons, this was not possible. Instead, clinical judgment, refined by a consensus process, was used to assess the expected prognosis in the two situations. In clinical practice, it is assumed that doctors can make this type of judgment about the patient in a consistently valid and rational manner. These judgments are important determinants of resource allocation in the health care system. In an investigation of the relation of health benefit to resource utilization, a method based on clinical judgment will therefore give meaningful results even in the absence of validation by external criteria.

Because the method involves assessments of hypothetical situations with varying health states, patient self-assessment of quality of life was not possible. Medical knowledge and experience are necessary to make these kinds of judgments. It can be argued that the patients could have been provided the necessary information by their doctors, but for practical reasons this would only have been possible for a small number of patients, and only for those well enough to participate. For some patients, it would have been unethical to provide detailed prognostic scenarios for the purpose of this study alone.

Time Trade-off Method

In the study reported here, time trade-off assessments were made using marks on a time scale instead of interviews. A similar technique was used by Pliskin et al. (12) in a questionnaire in which they let judges directly assess the number of years to trade-off for improvement in quality of life. This procedure gives the number of years in full health equivalent to the patient's life expectancy and quality of life, and can be regarded as a direct assessment of healthy years equivalents (HYE), a measure of health status proposed by Mehrez and Gafni (10). They argue against the use of time trade-off for measuring HYE as we have done, because the standard gamble technique must be used to place the HYE within the framework of utility theory (6). However, several authors argue convincingly that the methods are equivalent (2;4; 8;13).

In many applications of the time trade-off technique, a constant level of quality of life is assumed. In this study, most of the patients would be expected to have varying quality of life, and this assumption could not be made. When making time trade-off in this situation, the HYE of the patient's lifetime health profile was assessed directly. Mehrez and Gafni (10) also evaluated an entire lifetime health profile with varying health state directly, but used the standard gamble instead of the time trade-off technique.

To find the mean quality of life, we calculated the ratio between the HYE and the patient's total remaining lifetime. Since the HYE implicitly incorporates time preference, i.e., the tendency to value future health states lower than present ones, this is the mean quality of life after discounting future health states.

Composition of Expert Panels

Three specialists of internal medicine in each panel might have performed better than three different specialists, but we believe that this would have overemphasized the importance of the specialized professional viewpoint. This study was concerned with the final effect of the hospital stay on the patient's health and, therefore, a broader perspective than that provided by three internists was needed. The general practitioner has experience with long-term follow-up outside the hospital of many of the patients treated in departments of internal medicine and with patients with similar conditions who for various reasons are never admitted to hospital. The surgeon and the internist often cooperate closely and treat many of the same diseases. The level of agreement obtained indicates that the interaction between these three perspectives was useful.

Even though considerations about life expectancy and quality of life underlie decisions about patients in clinical practice, clinicians rarely evaluate these quantities numerically. The consensus process was essential for limiting the variation that could be expected when doctors were asked to do so. Nevertheless, there was only agreement

about broad categories of health gain. Given the paucity of empirical data and the degree to which the experts were left to rely on judgment alone, this was not surprising.

CONCLUSION

The method described here cannot be used by individual doctors for accurately assessing improvements in life expectancy and quality of life for individual patients, but was shown to produce reliable results when used by expert panels for identifying groups of patients with low, intermediate, and high health gain. The level of agreement was well above that expected from chance and better than that between most peer assessments in a recent review of other studies. Moreover, it was only slightly lower than the level of agreement for other generally accepted clinical methods.

APPENDIX

To illustrate the assessment technique, some cases with common conditions seen in a department of general medicine are presented below.

Agreement in both categories of life expectancy gain and quality of life gain:

Man, 77 years old, widower, retired farmer. Diabetes mellitus from 1976, treated with an oral agent. Terminated this medication himself after having experienced side effects. Admitted for initiation of insulin treatment. Symptoms and physical findings consistent with peripheral neuropathy. Serum glucose 30 mmol/L on admittance. Given two injections a day of intermediate acting insulin with resulting improvement in serum-glucose. Appointment made for further adjustment of insulin dose in the outpatient clinic.

Life expectancy gain:	panel A, 3.0 years	panel B, 0.6 years
Quality of life gain:	panel A, 0.03	panel B, 0.03

Agreement in category of life expectancy gain, disagreement in category of quality of life gain:

Man, 63 years old, retired fisherman. Except for musculoskeletal pain, not previously ill. Admitted with acute chest pain caused by an acute posterolateral myocardial infarction. Treated with streptokinase. Course complicated by transient clinical signs of pulmonary congestion, pneumonia, and a possible postmyocardial infarction syndrome. Echocardiography demonstrated pronounced hypokinesia of the posterolateral wall of the left ventricle and some pericardial effusion. Discharged with aspirin and enalapril.

Life expectancy gain:	panel A, 10.0 years	panel B, 7.0 years
Quality of life gain:	panel A, 0.45	panel B, 0.10

Disagreement in category of life expectancy gain, agreement in category of quality of life gain:

Woman, 67 years old, married, on sick leave from job as shop assistant. Hypertension. Angina pectoris for 1 year, NYHA class III despite treatment with propranolol, isosorbide dinitrate, and diltiazem. Admitted for percutaneous transluminal coronary angioplasty. Tandem stenosis in the second segment of the left anterior descending

artery dilated successfully. Treated with heparin for 1 day because of uncertainty about a possible intimal lesion, no signs of myocardial infarction. Performed 100W on exercise ECG before discharge. Discharged with reduced doses of propranolol and diltiazem plus aspirin.

Life expectancy gain:	panel A, 0 years	panel B, 2.0 years
Quality of life gain:	panel A, 0.13	panel B, 0.13

Agreement on no gain for both measures:

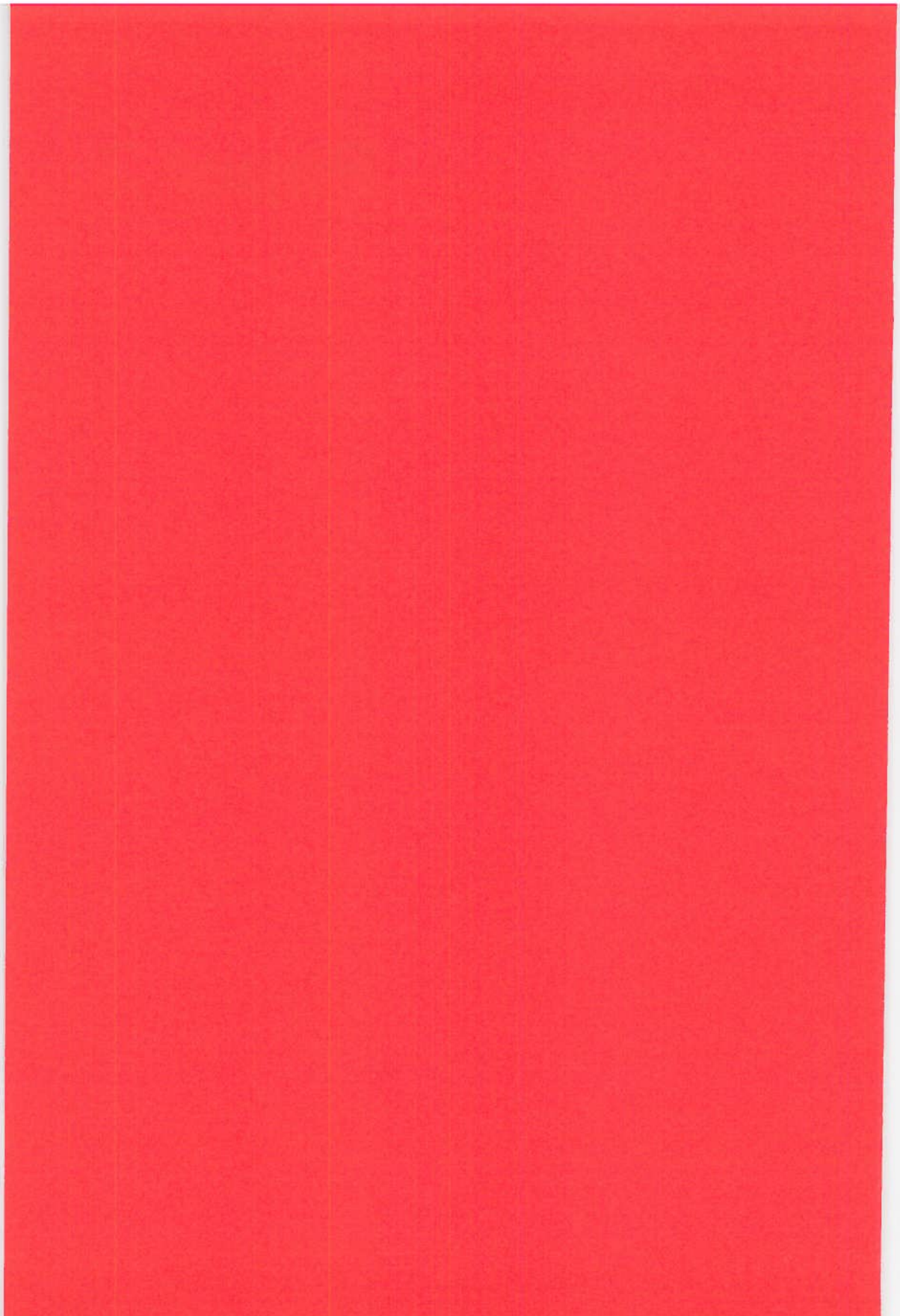
Man 68 years, retired fisherman. Several stays for chronic obstructive pulmonary disease, acute myocardial infarction 5 years ago. Admitted for worsening of his dyspnea and acute chest pain. No evidence of new myocardial infarction. Treated with prednisolone for his chronic obstructive pulmonary disease. Doxycycline was added because he also had fever. Discharged after gradual improvement of his dyspnea.

Life expectancy gain:	panel A, 0 years	panel B, 0 years
Quality of life gain:	panel A, 0	panel B, 0

REFERENCES

1. Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. *Discrete multivariate analysis.* Cambridge, MA: The MIT Press, 1975.
2. Buckingham, K. A note on HYE (healthy years equivalent). *Journal of Health Economics,* 1993, 11, 301–09.
3. Cohen, J. Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin,* 1968, 70, 213–20.
4. Culyer, A. J., & Wagstaff, A. QALYs versus HYEs. *Journal of Health Economics,* 1993, 11, 311–23.
5. Fleiss, J. L. *Statistical methods of rates and proportions,* 2nd ed. New York: John Wiley & Sons Inc., 1981.
6. Gafni, A., Birch, S., & Mehrez, A. Economics, health and health economics: HYEs versus QALYs. *Journal of Health Economics,* 11, 325–39.
7. Goldman, R. L. The reliability of peer assessments of quality of care. *JAMA,* 1992, 267, 958–60.
8. Johannesson, M., Pliskin, J. S., & Weinstein, M. C. Are healthy-years equivalents an improvement over quality-adjusted life years? *Medical Decision Making,* 1993, 13, 281–86.
9. Koran, L. M. The reliability of clinical methods, data and judgments. *New England Journal of Medicine,* 1975, 293, 642–46.
10. Mehrez, A., & Gafni, A. The healthy-years equivalents: How to measure them using the standard gamble approach. *Medical Decision Making,* 1991, 11, 140–46.
11. Nord, E. The QALY: A measure of social value rather than individual utility? *Health Economics,* 1994, 3, 89–93.
12. Pliskin, J. S., Shepard, D. S., & Weinstein, M. C. Utility functions for life years and health status. *Operations Research,* 1980, 28, 206–24.
13. Richardson, J. Cost utility analysis: What should be measured—Utility, value, or health equivalents? *Social Science and Medicine,* 1994, 39, 7–21.
14. Tanner, M. A., & Young, M. A. Modeling agreement among raters. *Journal of the American Statistical Association,* 1985, 80, 175–80.
15. Torrance, G. W. Health status index models: A unified mathematical view. *Management Science,* 1976, 22, 990–1001.
16. Torrance, G. W. Measurement of health state utilities for economic appraisal. *Journal of Health Economics,* 1986, 5, 1–30.

Paper 2



Clin Epidemiol Vol. 50, No. 9, pp. 987–995, 1997

ELSEVIER

0895-4356/97/$17.00
PII S0895-4356(97)00120-0

Does Admission to a Medical Department Improve Patient Life Expectancy?

Bjørn O. Eriksen,[1,] Ivar S. Kristiansen,[2] Erik Nord,[3] Jan Fr. Pape,[1] Sven M. Almdahl,[4] Anne Hensrud,[5] Steinar Jæger,[6] Fred A. Mürer,[7] Reidar Robertsen,[8] and Glen Thorsen[9]*

[1]Department of Medicine, University Hospital of Tromsø, 9038 Tromsø, Norway; [2]Institute of Community Medicine, University of Tromsø, 9017 Tromsø, Norway; [3]National Institute of Public Health, 0462 Oslo, Norway; [4]Department of Surgery, University Hospital of Tromsø, 9038 Tromsø, Norway; [5]Kommunelegekontort I Bardu, 9250 Bardu, Norway; [6]Department of Medicine, Norland Central Hospital, 8000 Bodo, Norway; [7]Department of Medicine, Rana Hospital, 8601 Mo, Norway; [8]Åsgård Psychiatric Hospital, 9017 Tromsø, Norway; and [9]Department of Surgery, Harstad Sykehus, 9400 Harstad, Norway

ABSTRACT. Doubts about the effectiveness of medical care in improving patient health have been raised by epidemiological studies and by studies of geographical variation and inappropriate use of health care. To investigate this problem, the life expectancy gain (LEG) from consecutive admissions to a department of internal medicine during a six-week period was assessed by two expert panels, each consisting of an internist, a surgeon, and a general practitioner. The mean LEG for all admissions was 2.25 years (n = 422). Sixty-one percent had a LEG of 0.10 years or less, while 5% had a LEG of more than 9.98 years. In a probabilistic sensitivity analysis, the mean LEG remained greater than zero under assumptions of overestimated positive LEG and underestimated negative LEG. We conclude that the life expectancy of the majority of the patients was not influenced by the admission, but that a minority had substantial gains, resulting in a high overall mean LEG. J CLIN EPIDEMIOL 50;9:987–995, 1997. © 1997 Elsevier Science Inc.

KEY WORDS. Life expectancy, outcome assessment health care, quality of care, Monte Carlo method, sensitivity analysis, iatrogenic disease.

INTRODUCTION

Despite impressive medical triumphs over the last decades, health care has come under attack, and the scientific foundation of medical practice is being questioned [1,2]. Claims can be heard that medical care has little [3,4] or even a negative effect on population health [5,6], and that scarce resources are being used inefficiently [7,8]. Studies of appropriateness of care and of practice variation indicate that all health care cannot be equally effective [9,10]. New technologies are often introduced without proper scientific evaluation [11,12], while randomized clinical trials sometimes show that well-established technologies yield no health benefit when they are evaluated in the end [13,14]. Also, the decline in mortality from infectious diseases, prior to the introduction of immunization and antimicrobial agents, indicates that medical innovations may have been less important contributors to health improvements in this century than is sometimes believed [15–17].

Address for correspondence: Bjørn Odvar Eriksen, Department of Medicine, University Hospital of Tromsø, 9038 Tromsø, Norway.
Accepted for publication on 9 June 1997.

Though none of the studies referred to above directly investigates the benefit obtained by individual patients from encounters with the health-care system, they all suggest that on average it may be low or even non-existent. The aim of the present investigation, which was undertaken as a part of the Tromsø Medical Department Health Benefit Study [18], was to explore this possibility by assessing the gain in life expectancy from consecutive admissions to the department of internal medicine of a university hospital. To investigate claims of inefficiency, we were particularly interested in the proportion of admissions with no or very low life expectancy gain. Ideally, estimation of life expectancy gain should be based on the results of randomized clinical trials (RCTs). However, a recent study found that only 53% of the primary interventions applied to patients in a department of general medicine were supported by RCTs [19]. In addition, the external validity of RCTs can sometimes be questioned because they are performed on selected patient groups and often cannot be applied directly to other patients. Thus, estimation of the life expectancy gain from hospital stays from this kind of "hard" evidence alone is not possible at present. As the second best solution, we chose

a method where life expectancy gain was assessed by panels of expert clinicians. This method has been shown to produce reliable results for a random sample of the admissions included in the Tromsø Medical Department Health Benefit Study [18], and has also been used in other similar studies [20]. However, a method based on clinical judgment has its obvious limitations. For this reason, the robustness of our conclusions was tested in a sensitivity analysis assuming different degrees of bias in the assessments. In particular, data from the literature about the occurrence of adverse events during hospitalization were used to investigate the effect of a possible underestimation of iatrogenic life expectancy loss.

MATERIAL AND METHODS

Subjects

In 1993, 5151 patients were admitted to the department of internal medicine at the University Hospital of Tromsø in the northern part of Norway. During a six-week period from 1 February 1993, all admissions were eligible for inclusion in the Tromsø Medical Department Health Benefit Study. Patients transferred from other university hospitals ($n = 3$), patients admitted for evaluation or continuation of treatment started during a previous stay ($n = 27$), and patients admitted for inclusion in drug trials ($n = 2$) were excluded, as well as one patient whose medical record could not be found. Nine planned readmissions were merged with the primary admission, resulting in a total of 479 included admissions. For a study of interpanel agreement, a random sample was obtained by giving each admission a probability of 0.1 of being drawn. The results of this study have been published previously [18]. The remaining admissions were used for the present investigation.

The study was approved by the Regional Ethics Committee and the Norwegian Data Inspectorate.

Expert Panels

Two expert panels were recruited, each consisting of an internist, a surgeon, and a general practitioner. All the experts were board-certified specialists in their respective fields. None of them had any connection with the department being investigated.

Assessment of Life Expectancy Gain (LEG)

When a patient was discharged or died in the hospital, a summary containing his complete medical history and all data from the current stay was compiled by the project coordinator, a board-certified specialist of internal medicine. The summaries were intended to be comprehensive, and included a social history, previous illnesses, current health problem, medication, physical findings, results of tests, diagnosis, treatment during the stay, and plans for further treatment. They were used by the experts for assessing various aspects of health benefit from the hospital stays. The result of the evaluation of life expectancy gain (LEG) will be reported here.

To estimate the gain in life expectancy attributable to the hospital stay, the experts estimated life expectancy for two situations: (i) for the patient's prognosis after this hospital stay, taking into account the expected outcome of planned treatment after discharge, and (ii) for the patient's expected prognosis in the hypothetical situation had he not been admitted to hospital or treated elsewhere for his current health problem. LEG was then calculated as the difference between these two assessments. The experts were instructed to base their assessments on the best available evidence in each case: RCTs, other empirical data, or clinical judgment alone. They were also told to consider the influence of other diseases and risk factors on life expectancy. As an aid, the experts were given information about the average life expectancy of a person of the same sex and age in the general population.

The experts also assessed whether patients with a positive LEG could have achieved the same gain in an outpatient clinic or in primary care.

Assessment Protocol

Each admission was randomly assigned to be assessed by one of the two expert panels. In the panels, the admissions were first assessed by each expert individually. The estimates of the three members of each panel were then compared. Consensus was defined to exist when the difference between the maximum and minimum LEG estimates did not exceed 25% of the average estimated life expectancy of the patient after the hospital stay. When this criterion was met, the panel's assessment was defined as the median of the three individual assessments. Otherwise, the case was discussed in a meeting of the three members of the panel. After the discussion, the experts revised their individual estimates, and the median was again taken as the LEG, even if the consensus criteria were not met.

There was no contact between the two expert panels during the study.

ICD9 Codes

All ICD9 codes were truncated to three digits and checked by the project coordinator for consistency with the diagnostic conclusions in the discharge reports. When there was more than one code, he also checked that the principal diagnosis corresponded to the patient's current health problem.

Statistical Methods

Approximate 95% confidence intervals of statistical parameters were estimated by taking the 2.5th and 97.5th percen

tiles of the bootstrap distribution of the parameter in question. The bootstrap distributions were obtained with Monte Carlo simulations by drawing 1000 random resamples of size 422 with replacement from the original observations. The bootstrap distributions of regression coefficients in multivariate linear regression analyses were found by calculating the least-squares estimates of the coefficients for each of 10,000 resamples.

Sensitivity Analysis

The mean LEG for all admissions is a function of the proportion of admissions achieving LEG and the magnitude of the LEG obtained through each admission. From this amount must be subtracted iatrogenic life expectancy losses (i.e., negative LEG), which are a function of the proportion of admissions suffering loss and the magnitude of loss suffered by each admission. To investigate the dependence of the mean LEG on these four variables, a probabilistic sensitivity analysis was performed [21]. Following a method described by Doubilet *et al.* [22], the variables were varied simultaneously by drawing them from logistic-normal probability density distributions in a Monte Carlo simulation. In a logistic-normal distribution, the logit transform-log(X/1-X)-, of each variable is normally distributed. For each variable, the parameters of this distribution were calculated from the baseline value and the bounds of a 95% confidence interval.

The baseline proportion of admissions obtaining a positive LEG was taken from the present study, and the lower and upper bounds of this variable were set equal to the estimates of expert panels A and B, respectively. The baseline magnitude of LEG and its 95% confidence interval were also estimated on the basis of our own data by calculating the mean LEG for admissions with LEG greater than 0.10 year.

Estimates of the proportion of admissions resulting in life expectancy loss were found in the literature. The percentage of patients suffering an iatrogenic death in departments of internal medicine was estimated by Kneel at 2% [23] and by Brennan at 0.5% [24]. The percentage suffering major adverse events, defined as events that produce considerable disability or threaten life, was 9% in Kneel's study, while the percentage with permanent disability was 0.1% in Brennan's study. The sums of the two estimates for each of the studies were taken as the lower (0.6%) and upper (11%) bounds for the percentage of admissions with negative LEG, and their average as the baseline percentage.

The baseline amount of negative LEG suffered by these admissions was arbitrarily set at 50% of the average life expectancy of a person in the general population who is of the same age and sex as the patient. The lower and upper bounds were set at 25% and 75%.

The analysis was repeated with the additional assumption that all LEGs were overestimated by 50%.

RESULTS

Of the 422 patients included in the study, 160 (37.9%) were women, and 262 (62.1%) men. The mean age was 61.6 years; for women 61.0 years (range 16–94), for men 61.9 years (range 15–90). 152 (36.0%) were elective and 270 (64.0%) emergency admissions. Twenty (4.7%) patients died in the hospital.

Diagnosis

In total, 110 different ICD9 principal diagnosis were used. Similar diagnoses were merged so that each diagnostic group included 10 hospital stays or more (Table 1). Angina pectoris and acute myocardial infarction together accounted for 27.2% of the admissions.

Differences Between the Two Expert Panels

Two hundred fifteen admissions were assessed by expert panel A (50.9%), and 207 by expert panel B (49.1%). The difference between the mean LEG of these two groups was 0.32 years (95% confidence interval −0.88–1.42). The percentage of admissions assessed to have had a gain less than 0.10 year was 70.2% by panel A and 52.2% by panel B. The difference between the two was 18.0% (95% confidence interval 8.7–26.9%). In the following analyses, the estimates of the two panels were pooled.

Life Expectancy Gain (LEG)

The total LEG for all admissions was 949.17 years, and the mean LEG 2.25 years (95% confidence interval 1.74–2.84). Only one stay (0.2%) was estimated to have resulted in a negative LEG, i.e., that the hospital stay shortened the patient's life. This patient was an 80-year-old man who had initially been admitted for hematochezia, and who died after surgery for a suspected sigmoid cancer. His LEG was −0.07 years, which is a life expectancy loss of about 1 month. The final diagnosis was diverticulitis with obstruction, which probably also would have been fatal if it had not been treated surgically.

Of the admissions, 259 (61.4%) had a LEG of 0.10 years or less, while 5% had a LEG of more than 9.98 years. The distribution of LEG is shown in Fig. 1, and the LEG according to sex, age group, admission category, and diagnostic group in Table 1. The assessments for the 10 patients with the highest LEG are presented in Table 2. These patients together accounted for 33.1% of the total LEG in the material.

Regression Analysis

The effects of sex, age, diagnosis, and admission type (elective or emergency) on LEG were examined in a multivariate linear regression analysis. Dummy variables were used for

TABLE 1. Mean LEG according to sex, age group, admission category and diagnostic group for patients admitted to a department of internal medicine (n = 422)

	ICD9-code	n (%)	Mean LEG in years (95% CI)	Percent of total LEG
Total	—	422 (100.0)	2.25 (1.74–2.84)	100.0
Sex				
Men	—	262 (62.1)	2.03 (1.56–2.59)	56.1
Women	—	160 (37.9)	2.60 (1.51–3.96)	43.9
Age group				
<50 years	—	93 (22.0)	4.12 (2.19–6.49)	40.4
50–69 years	—	180 (42.7)	2.18 (1.61–2.80)	41.4
≥70 years	—	149 (35.3)	1.16 (0.88–1.45)	18.2
Admission category				
Elective	—	152 (36.0)	1.81 (1.31–2.37)	29.0
Emergency	—	270 (64.0)	2.50 (1.73–3.38)	71.0
Diagnostic group				
Infectious diseases	001–139	17 (4.0)	8.88 (1.25–18.86)	15.9
Malignant diseases	140–208	42 (10.0)	0.95 (0.58–1.38)	4.2
Endocrinological diseases	240–259	11 (2.6)	12.28 (4.36–21.17)	14.2
Acute myocardial infarction	410	30 (7.1)	1.03 (0.35–1.83)	3.3
Angina pectoris	411–414	85 (20.1)	1.79 (1.15–2.53)	16.0
Other heart diseases	420–429	45 (10.7)	2.63 (1.78–3.50)	12.4
Cerebrovascular diseases	430–438	21 (5.0)	0.22 (0.00–0.49)	0.5
Pneumonia and influenza	480–487	16 (3.8)	2.97 (1.38–5.02)	5.0
Chronic obstr. pulm. disease	496	20 (4.7)	1.24 (0.10–2.99)	2.6
Hepatobiliary/pancreatic diseases	570–579	13 (3.1)	2.23 (0.22–4.98)	3.0
Undiagnosed symptoms	780–789	30 (7.1)	0.07 (0.00–0.23)	0.2
Other		92 (21.8)	2.33 (1.40–3.55)	22.6

Abbreviations: CI = confidence interval, LEG = life expectancy gain.

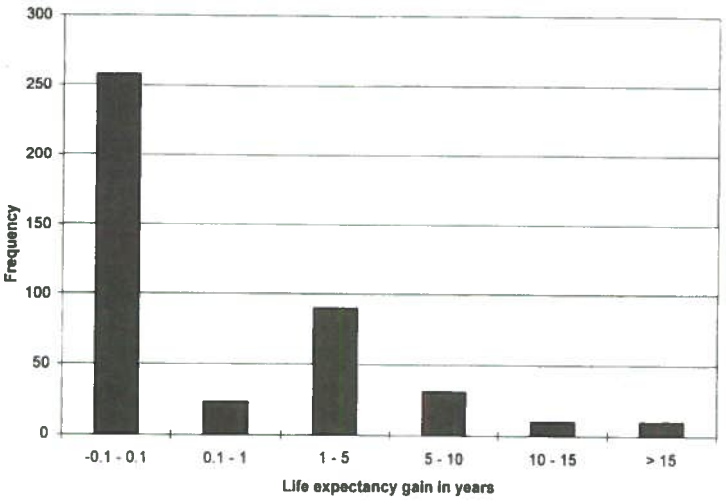


FIGURE 1. Distribution of LEG from hospital stays as assessed by the two expert panels (n = 422).

TABLE 2. The ten admissions with the highest LEGs from stays in a department of internal medicine ($n = 422$)

Sex	Age (years)	ICD9 code	Admission category	Mean remaining lifetime in general population (years)	Remaining lifetime		Life expectancy gain[b] (years)	Clinical details
					Without admission[a]	After this admission[a] (years)		
Female	18	036	Emergency	63	4 days	63	63	Treatment for meningococcal septicemia
Female	28	038	Emergency	53	5 days	51	51	Treatment for septicemia caused by group A streptococci
Male	24	250	Emergency	50	2 days	40	40	Patient with known insulin-dependent diabetes mellitus treated for ketoacidosis
Female	23	790	Emergency	58	7 days	40	40	Relapsing bacteremias treated with antimicrobial prophylaxis
Female	50	250	Emergency	32	5 years	27	25	Treatment for newly diagnosed insulin-dependent diabetes mellitus
Male	47	242	Emergency	29	1 year	25	24	Treatment for thyreotoxic cardiomyopathia
Female	52	250	Emergency	30	5 days	24	24	Treatment for diabetic ketoacidosis and pulmonary abscess
Female	42	710	Elective	39	13 years	30	17	Systemic lupus erythematosus with stenosis of the aortic valve prepared for surgical treatment
Male	45	413	Elective	30	10 years	25	15	Proximal stenosis of left anterior descending coronary artery treated with percutaneous transluminal coronary angioplasty
Male	50	413	Emergency	26	5 years	20	15	Aortocoronary bypass for occluded left anterior descending coronary artery and occluded right coronary artery after previously being resuscitated for ventricular fibrillation

[a] Median of the assessments of the three members of the expert panel.
[b] Median of the assessments of the three members of the panel, thus not necessarily equal to the difference between the values in the two preceding columns.

TABLE 3. Multivariate linear regression analysis of LEG (n = 422)

Independent variables[a]	Estimate	95% CI[b]
Intercept	5.79	(3.05–9.21)
Sex (0 = M, 1 = F)	0.09	(−0.95–1.25)
Age	−0.07	(−0.13–−0.02)
Admission category (0 = elective, 1 = emergency)	0.88	(−0.15–2.01)
Infectious diseases	5.21	(−1.63–15.01)
Malignant diseases	−0.60	(−1.74–0.53)
Endocrinological diseases	9.85	(2.38–18.09)
Acute myocardial infarction	−0.74	(−2.04–0.57)
Angina pectoris	0.13	(−1.15–1.40)
Other heart diseases	1.21	(−0.15–2.69)
Cerebrovascular diseases	−1.09	(−2.18–0.05)
Pneumonia and influenza	1.42	(−0.46–3.50)
Chronic obstructive pulmonary disease	−0.76	(−2.46–1.33)
Hepatobiliary/pancreatic diseases	0.21	(−2.26–3.20)
Undiagnosed symptoms	−2.06	(−3.35–−1.06)

Abbreviations: CI = confidence interval, LEG = life expectancy gain.
[a]The disease category "other" serves as reference for the dummy variables of the disease categories.
[b]Estimated with the bootstrap algorithm from 10,000 resamples.

the diagnostic groups with "other diagnoses" as reference. Because of non-normal residuals, the bootstrap algorithm was used for finding 95% confidence intervals for the regression coefficients. The confidence intervals of the coefficients for age, endocrinological diseases, and undiagnosed symptoms did not include zero. Higher age and undiagnosed symptoms were associated with lower and endocrinological diseases with higher LEG (Table 3).

Level of Care

Five of the patients could have obtained a similar LEG in primary care or in an outpatient clinic. The total LEG of these patients was 9.04 years, which was 1.0% of the total LEG in the material.

Probabilistic Sensitivity Analysis

The baseline and lower and upper bounds for the variables in the probabilistic sensitivity analysis model are shown in Table 4. In a Monte Carlo simulation of 10,000 runs of the model, the distribution of the mean LEG had a median 1.40 years (mean 1.34, standard deviation 0.42, ran −1.48–2.57 years, 2.5th percentile 0.36 years, 97.5th pe centile 2.04 years). A total of 99.2% of the runs resulted a mean LEG greater than zero.

Running the model under the additional assumption th all positive LEGs had been over-estimated by 50% result in a median mean LEG of 0.76 years (mean 0.71, standa deviation 0.36, range −2.13–1.60 years, 2.5th percent −0.14 years, 97.5th percentile 1.27 years). A total of 95.9 of the runs yielded a mean LEG greater than zero.

TABLE 4. Data used for probabilistic sensitivity analysis of mean LEG

	Baseline (mean)	Lower bound	Upper bound
Probabilities			
Positive life expectancy gain	0.388	0.298	0.478
Negative life expectancy gain	0.058	0.006	0.110
Life expectancy gain			
Positive life expectancy gain (years)	5.82	5.03	6.62
Negative life expectancy gain, (fraction of life expectancy in general population)	0.50	0.25	0.75

DISCUSSION

Prolongation of life is one of the primary aims of heal care. The degree to which this aim is attained in routi clinical practice is obviously of great interest to clinicia health administrators, and politicians. The present inves gation has addressed this issue by focusing on internal me cine, which accounts for a large part of patient care hospitals.

When studying the LEG from unselected admissions a department of internal medicine, assessment by exp panel is probably the best method available at present. a previous study of the reliability of such assessments wh categorized as low, intermediate, and high LEG, we report an overall agreement of 0.67 and a weighted kappa of 0. [18]. This level of agreement is usually regarded as "fair good" [25]. However, though reliable, the assessments m all have been subject to the same bias [26]. To avoid so of the most obvious sources of bias, we chose experts w had no connection with the department being studi Also, surgeons and general practitioners were included the panels at least in part because it was assumed that tl

vould be less susceptible to upward bias than would inter-
ists.

In other studies using expert panels, specific guidelines
or evaluating various outcomes have often been made from
iterature studies and expert opinion. In our study, it was
ot feasible to use this method for all the different cases
dmitted to a department of internal medicine. Instead, the
xperts were instructed to use the best evidence available
n each case. They were also instructed to take into consid-
ration all relevant aspects of the patient's situation that
ight influence his life expectancy, including other ill-
esses and risk factors.

Mean LEG

Our main finding was that mean LEG from admissions to department of internal medicine was 2.25 years, which clearly does not support the claim that medical care has little or no positive effect on patients' health.

A probabilistic sensitivity analysis was used to investigate the effect of possible bias on the conclusion that mean LEG was greater than zero. We assumed that upward bias could result from (i) awarding LEG to patients who actually did not benefit, and (ii) underestimating a possible negative effect of the hospital care on patient life expectancy, i.e., that some patients had actually suffered iatrogenic life expectancy loss. In the model, both the percentages of admissions resulting in positive and negative LEG, as well as the magnitudes of positive and negative gain, were varied simultaneously in a Monte Carlo simulation.

We were especially interested in studying the effect of a higher percentage of iatrogenic life expectancy loss than that estimated by the expert panels (0.2%). Therefore, baseline data for this effect were taken from the literature. As far as we know, studies of the occurrence of life expectancy loss as such do not exist, but at least two studies provide estimates of the probability of major adverse events, these being defined as events that produce considerable disability or threaten life [23,24]. The estimates differ widely (0.6% versus 11%), and the baseline probability of life expectancy loss was taken as their average (5.8%). This percentage was nearly 30 times the estimate of the expert panels (0.2%). We could not find information on the magnitude of life expectancy loss suffered by each patient in the literature, and the baseline of this variable was arbitrarily set as high as 50% of the remaining life expectancy of a person of the same age and sex in the general population. Since patients have a shorter life expectancy than the general population, there is good reason to believe that the true value is lower, which means that this assumption would bias the model toward a lower mean LEG.

When running the model with these inputs (Table 4), the median mean LEG was lower than estimated by the expert panels, but still as high as 1.40 years. Ninety-nine percent of the 10,000 Monte Carlo simulations gave a positive mean LEG. Repeating the run under the added assumption that all positive LEGs had been overestimated by 50% still resulted in a positive median mean LEG of 0.76 years, and a positive mean LEG in 96% of the simulations.

It is not clear whether these results from the medical department of a teaching hospital are representative of other parts of health care. One would expect better results from a teaching hospital than from a local hospital, but we are not aware of data that would support this assumption. In surgical departments, the opportunities both for life expectancy gains and losses may be greater than in departments of internal medicine, but studies would be needed to find out whether the balance is positive or negative.

The finding of a zero or negative mean LEG would have supported the claim that medical care has little effect on population longevity, but a positive mean LEG does not necessarily imply a positive effect on the population level. An estimate of the effect on the population level would also have to take into account the proportion of the population treated in hospitals and the frequency of readmissions. For example, if a patient is saved from diabetic coma several times, this adds to population life expectancy only once, but each admission would increase the mean LEG of a hospital department.

However, even if our results do not directly contradict the views of the most extreme critics of health care [5,6], other studies, which have examined the effect of medical care on the population level, do. Mackenbach *et al.* [27] found the gain in life expectancy in the Netherlands from the 1950s to 1980s due to the reduction in mortality for conditions amenable to medical treatment to be 2.96 years for males and 3.95 years for females. Studies of causes of death amenable to medical treatment in other countries show similar results [28–30]. Bunker *et al.* [31] estimated the effect of curative medicine for selected diagnoses on life expectancy at birth in the United States from data about the effect of treatments and population at risk. He found that curative medicine prolonged life expectancy about 3.5 years. Hadley's [32] study of mortality rates in the United States also concluded with a significant effect of medical care. These studies all indicate a non-trivial effect of medical care on life expectancy.

Distribution of LEG

Although the mean LEG was positive, the hospital stays had little or no influence on the life expectancy of the majority of the patients. On the other hand, a minority had considerable benefits (Fig. 1). The positive mean LEG was a result of the high gain for these few patients, who, in the opinion of the expert panels, would have suffered premature deaths if they had not been admitted.

Of the ten patients with the highest gains, none was older

than 50 years (Table 2), and higher age was associated with a lower LEG in the regression analysis. Gender was not a significant regressor. Endocrinological diseases predisposed for high LEG, whereas patients who were not given a specific diagnosis were least likely to benefit (Table 3). The list of the ten patients with the highest gains (Table 2) shows that a large proportion of the total LEG came from treating life-threatening bacterial infections, complications of diabetes mellitus, and one patient with cardiac complications of thyrotoxicosis. Some patients with coronary heart disease also achieved high gains, but acute myocardial infarction and angina pectoris were not significantly associated with a high LEG (Table 3).

It is noteworthy that a considerable percentage of the total LEG was attributable to interventions that have been available for decades (hormone substitution, antimicrobial agents). Infectious and endocrinological diseases together accounted for 7% of the admissions and 30% of the total LEG. In most industrialized countries, these diseases are not very frequent causes of death. In contrast, malignant, and cardiovascular diseases were the cause of 53% of the admissions, but only 37% of the total LEG (Table 1).

A high proportion of admissions with low gain is consistent with the high rate of inappropriately performed procedures found in some studies [33]. An inappropriately performed procedure would have a low probability of a LEG but exposes the patient to an unnecessary risk of iatrogenic health loss. The sensitivity analysis indicated that there was a wide margin before this could outweigh a positive LEG, but the analysis did not consider loss in quality of life, which is probably more common than loss in life expectancy.

Geographical variation in the rate of hospital admissions without any noticeable difference in mortality [34] could be explained by variation in the number of admissions with a low LEG. As long as the minority of patients with high gain is identified and admitted, the total number of admissions will not necessarily be correlated with mortality.

Life Expectancy versus Quality of Life

It would be premature to conclude that the large percentage of admissions with negligible LEG were unnecessary or indicated inefficient use of health-care resources. Some claim that the effect of modern health care should be judged more from its effect on quality of life than on longevity [35–37]. It is possible that the majority of patients with a low LEG in this study had had an improvement in their quality of life, and that the percentage of patients with no benefit at all was lower. This issue will be addressed in another paper.

CONCLUSIONS

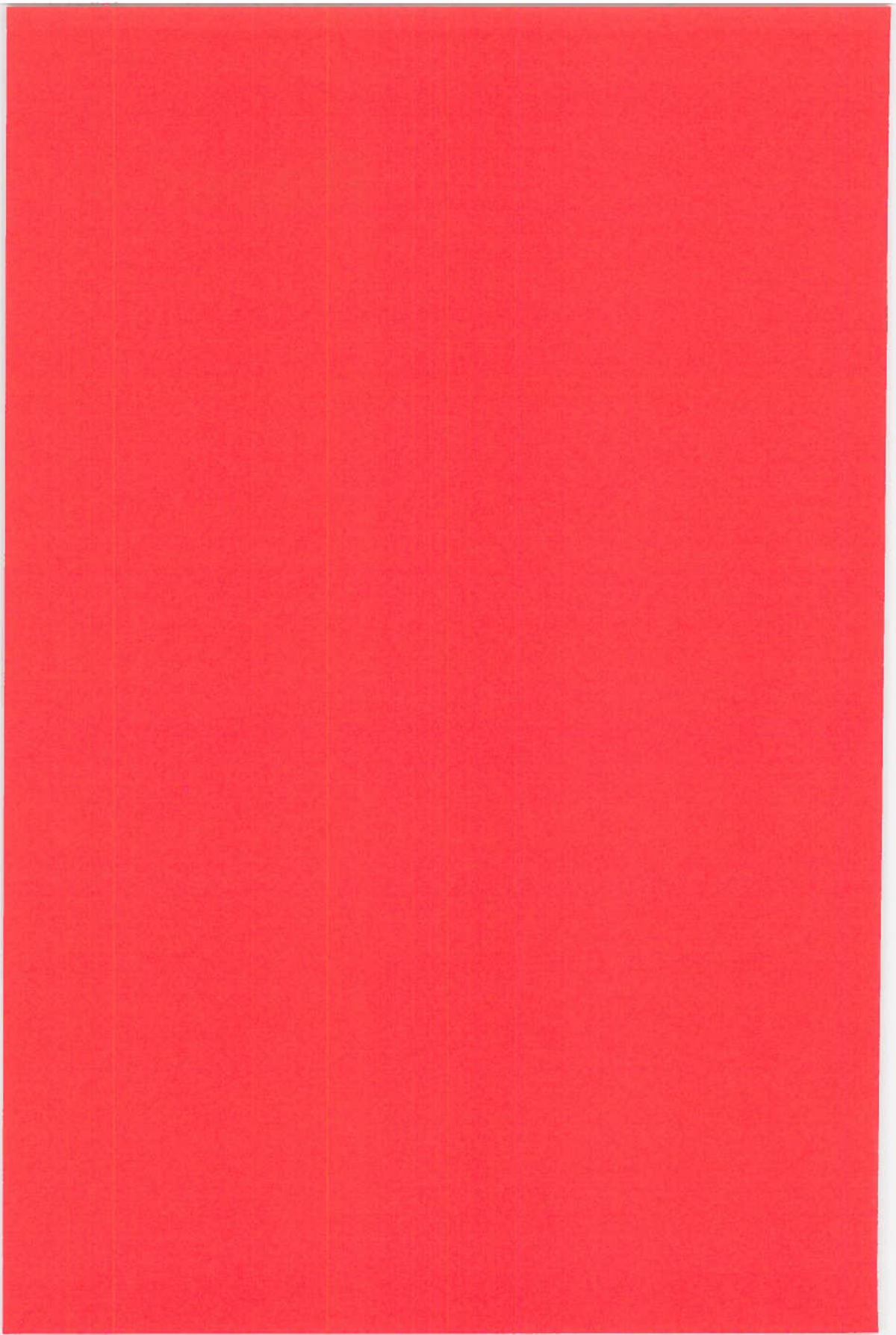
Admission to a department of internal medicine had no influence on the life expectancy of the majority of the patients. A minority had substantial LEGs, resulting in an overall mean LEG of 2.25 years. When assuming that LEG had been overestimated and iatrogenic life expectancy loss underestimated in a sensitivity analysis, the mean LEG was still positive in almost all 10,000 runs of a Monte Carlo simulation.

References

1. Roper WL, Winkenwerder W, Hackbarth GM, Krakauer H. Effectiveness in health care. An initiative to evaluate and improve medical practice. **N Engl J Med** 1988; 319: 1197–1202.
2. Eddy DM. The challenge. **J Am Med Assoc** 1990; 263: 287–290.
3. Evans RG, Barer ML, Marmor TR, Eds. **Why Are Some People Healthy and Others Not?** New York: Walter de Gruyter 1994.
4. Knowles JH. **Doing Better and Feeling Worse. Health in the United States.** New York: W.W. Norton & Co., Inc.; 1977.
5. Illich I. **Medical Nemesis: The Expropriation of Health.** New York: Random House; 1976.
6. Carlson RJ. **The End of Medicine.** 1st ed. New York: John Wiley & Sons; 1975.
7. Culyer AJ, Maynard AK, Posnett JW. **Competition in Health Care: Reforming the NHS.** London: MacMillan; 1990.
8. OECD. **Health Care Systems in Transition: The Search for Efficiency.** Paris: OECD; 1990.
9. Brook RH. Appropriateness: The next frontier. **Br Med J** 1994; 308: 218–219.
10. Wennberg J, Gittelsohn A. Small area variations in health care delivery. **Science (Wash DC)** 1973; 182 :1102–1108.
11. Banta HD, Luce BR. **Health Care Technology and Its Assessment: An International Perspective.** Oxford: Oxford University Press; 1993.
12. Grimes DA. Technology follies. The uncritical acceptance of medical innovation. **J Am Med Assoc** 1993; 269: 3030–3033.
13. Duffy SQ, Farley DE. The protracted demise of medical technology: The case of intermittent positive pressure breathing. **Med Care** 1992; 30: 718–736.
14. Freeman R. Intrapartum fetal monitoring—A disappointing story. **N Engl J Med** 1990; 322: 624–626.
15. McKinlay JB, McKinlay SM. The questionable contribution of medical measures to the decline of mortality in the United States in the twentieth century. **Milbank Mem Fund Q Health Soc** 1977; 55: 405–428.
16. Powles J. The effect of health services on adult male mortality in relation to the effects of social and economic factors. **Ethic Sci Med** 1978; 5: 1–13.
17. McKeown T. **The Role of Medicine. Dream, Mirage or Nemesis?** 2nd ed. Oxford: Basil Blackwell; 1979.
18. Eriksen BO, Almdahl SM, Hensrud A, Jæger S, Kristiansen IS, Mürer FA, *et al.* Assessing health benefit from hospitalization: Agreement between expert panels. **Int J Technol Assess Health Care** 1996; 12: 126–135.
19. Ellis J, Mulligan I, Rowe J, Sackett DL. Inpatient general medicine is evidence based. **Lancet** 1995; 346: 407–410.
20. Hotvedt R, Kristiansen IS, Forde OH, Thoner J, Almdahl SM, Bjorsvik G, *et al.* Which groups of patients benefit from helicopter evacuation? **Lancet** 1996; 347: 1362–1366.
21. Critchfield GC, Willard KE, Connelly DP. Probabilistic sensitivity analysis methods for general decision models. **Comput Biomed Res** 1986; 19: 254–265.
22. Doubilet P, Begg CB, Weinstein MC, Braun P, McNeil BJ. Probabilistic sensitivity analysis using Monte Carlo simulation. A practical approach. **Med Decis Making** 1985; 5: 157–177.

. Steel K, Gertman PM, Crescenzi C, Anderson J. Iatrogenic illness on a general medical service at a university hospital. N Engl J Med 1981; 304: 638–642.
. Brennan TA, Leape LL, Laird NM, Hebert L, Localio AR, Lawthers AG, *et al.* Incidence of adverse events and negligence in hospitalized patients. N Engl J Med 1991; 324: 370–376.
. Fleiss JL. **Statistical Methods of Rates and Proportions.** 2nd ed. New York: John Wiley & Sons, Inc.; 1981.
. Tversky A, Kahneman D. Judgement under uncertainty: Heuristics and biases. **Science (Wash DC)** 1974; 185: 1124–1131.
. Mackenbach JP, Looman CW, Kunst AE, Habbema JD, van der Mass PJ. Post-1950 mortality trends and medical care: Gains in life expectancy due to declines in mortality from conditions amenable to medical intervention in The Netherlands. **Soc Sci Med** 1988; 27: 889–894.
. Charlton JR, Velez R. Some international comparisons of mortality amenable to medical intervention. **Br Med J Clin Res Ed** 1986; 292: 295–301.
Poikolainen K, Eskola J. The effect of health services on mortality: Decline in death rates from amenable and non-amenable causes in Finland, 1969–81. **Lancet** 1986; 1: 199–202.
Charlton JR, Hartley RM, Silver R, Holland WW. Geographical variation in mortality from conditions amenable to medical intervention in England and Wales. **Lancet** 1983; 1: 691–696.
31. Bunker JP, Frazier HS, Mosteller F. Improving health: Measuring effects of medical care. **Milbank Q** 1994; 72: 225–258.
32. Hadley J. **More Medical Care, Better Health? An Economic Analysis of Mortality Rates.** First ed. Washington D.C. The Urban Institute Press; 1982.
33. Brook RH, Park RE, Chassin MR, Solomon DH, Keesey J, Kosecoff J. Predicting the appropriate use of carotid endarterectomy, upper gastrointestinal endoscopy, and coronary angiography. **N Engl J Med** 1990; 323: 1173–1177.
34. Wennberg JE, Freeman JL, Shelton RM, Bubolz TA. Hospital use and mortality among Medicare beneficiaries in Boston and New Haven. **N Engl J Med** 1989; 321: 1168–1173.
35. McDermott W. Medicine: The public good and one's own. **Perspect Biol Med** 1978; 21: 167–187.
36. Levine S, Feldman JJ, Elinson J. Does Medical Care Do Any Good? In: Mechanic D, Ed. **Handbook of Health, Health Care and the Health Professions.** New York: Free Press; 1983: 394–404.
37. Anderson OW, Morrison EM. The Worth of Medical Care: A Critical Review. **Med Care Rev** 1989; 46: 121–155.

Paper 3



Journal of Internal Medicine 1998; 244: 397–404

Does admission to a department of internal medicine improve patients' quality of life?

B. O. ERIKSEN[1], I. S. KRISTIANSEN[2], E. NORD[1], J. F. PAPE[1], S. M. ALMDAHL[4], A. HENSRUD[5], S. JÆGER[6], F. A. MÜRER[7], R. ROBERTSEN[8] & G. THORSEN[9]

From the [1]Department of Medicine, University Hospital of Tromsø; [2]Institute of Community Medicine, University of Tromsø; [3]National Institute of Public Health, Oslo; [4]Department of Surgery, University Hospital of Tromsø; [5]Kommunelegekontoret i Bardu, Bardu; [6]Department of Medicine, Nordland Central Hospital, Bodø; [7]Department of Medicine, Rana Hospital, Mo; [8]Åsgård Psychiatric Hospital, Tromsø; and [9]Department of Surgery, Harstad sykehus, Harstad, Norway

Abstract. Eriksen BO, Kristiansen IS, Nord E, Pape JF, Almdahl SM, Hensrud A, Jæger S, Mürer FA, Robertsen R, Thorsen G (University Hospital of Tromsø; University of Tromsø; National Institute of Public Health, Oslo; Kommunelegekontoret i Bardu, Bardu; Nordland Central Hospital, Bodø; Rana Hospital, Mo; Åsgård Psychiatric Hospital, Tromsø; and Harstad sykehus, Harstad, Norway). Does admission to a department of internal medicine improve patients' quality of life? *J Intern Med* 1998; **244**: 397–404.

Objectives. The Tromsø Medical Department Health Benefit Study was designed to estimate health gains from admissions to a department of internal medicine. We have previously reported that the hospital stays had no effect on the life expectancy of 61% of the patients. However, it has been claimed that modern medicine has a greater effect on quality of life (QoL) than on life expectancy. The aim of the present study was to investigate this issue by estimating gains in QoL for patients admitted to a department of internal medicine.

Design. The time trade-off method (TTO) was used for assessing QoL gain from consecutive admissions during a 6-week period. The assessments were made by one of two expert panels, each consisting of an internist, a surgeon and a general practitioner, on the basis of summaries of all relevant clinical information about the patients. Short-term improvements in QoL during the stay or shortly after discharge were scored on an ordinal scale.

Results. Of the admitted patients, 41% had gains in QoL measured with the TTO (mean gain = 0.06; 95% confidence interval = 0.05–0.07; n = 422), and eight of these had gains equal to or greater than 0.50. Another 40% had gains in health-related short-term QoL measured with the ordinal scale. In a multivariate linear regression analysis, emergency admissions, high age and the disease categories 'endocrinological diseases' and 'pneumonia and influenza', were associated with higher gain, and 'undiagnosed symptoms' and 'cerebrovascular diseases' with lower gain.

Conclusions. As judged by the expert panels, the investigated department of internal medicine was effective in improving the QoL of 81% of the admitted patients. Whilst most of the patients achieved small gains, a minority had gains in QoL corresponding to the treatment of life-threatening diseases.

Keywords: health priorities, health services research, health status indicators, hospital, patient admission, quality of life.

Introduction

During the last 20 years, it has been discussed to what extent modern health care has a positive influence on the health of patients. Geographical variations in the use of health care [1–4] and a high proportion of inappropriately applied procedures in hospitals [5] suggest that all medical care cannot be equally effective. Furthermore, a high percentage of unnecessary admissions to hospitals [6–8] implies that many patients run the risk of complications from unnecessary interventions. In addition, epidemiological studies have cast doubt on the effect of health care from a population perspective [9–11].

The Tromso Medical Department Health Benefit Study was designed to assess gains in quantity and quality of health in a department of internal medicine [12]. We have previously reported that although some patients had substantial life expectancy gains, the life expectancy of as many as 61% was unaffected [13]. This could indicate a waste of resources, but some authors claim that modern health care should be judged more from its effect on quality of life than on life expectancy [14–16]. Accordingly, one should expect that more patients would have had improvements in quality of life than in life expectancy.

The aim of the present study was to address this issue by estimating gains in quality of life for the same patients as in our previous investigation of life expectancy gains [13]. To assess quality of life gain attributable to a hospital stay, it is necessary to make an explicit evaluation of the expected quality of life without hospital admission. Because this presumes medical knowledge, the assessments were made by panels of expert physicians. For measuring quality of life, we used the time trade-off technique (TTO), which has been validated by others [17] and has been found by us to produce reliable results for a random sample of the patients included in the Tromso Medical Department Health Benefit Study [12]. Methodological issues raised by the method have been discussed elsewhere [12].

Material and methods

The University Hospital of Tromso is located in the northern part of Norway. During a 6-week period starting on the 1 February 1993, all 521 admitted patients were considered for inclusion in the study. Patients transferred from other university hospitals ($n = 3$), admitted for evaluation or continuation of treatment started during a previous stay ($n = 27$) or admitted for inclusion in drug trials ($n = 2$) were excluded, as well as one patient whose medical record could not be found. Nine planned readmissions were merged with the primary admission, resulting in a total of 479 included admissions.

Two expert panels (A and B) were recruited, each consisting of an internist, a surgeon and a general practitioner. All the experts were board-certified specialists.

On admission, the 479 included admissions were randomized to group 1, 2 or 3 with probabilities of 0.10, 0.45 and 0.45, respectively. The patients in group 1 were all assessed by both expert panels for the purpose of investigating interpanel agreement. Group 2 was assessed by expert panel A only, and group 3 by panel B only. The experts were blinded to which admissions were included in group 1. The randomization resulted in 57 admissions in group 1, 215 in group 2 and 207 in group 3.

Assessments of gains in quality of life

When a patient was discharged or died in the hospital, a summary containing the complete medical history and all data from the current stay was compiled by the project coordinator (BOE), who is a board-certified specialist of internal medicine. The summary was used by the experts for assessing health benefits from the hospital stay. The results of the evaluation of gain in quality of life (QoL) will be reported here.

Long-term quality of life gain. The patients' expected QoL was assessed separately by the expert panels for two situations: (1) for the expected prognosis after the hospital stay, taking into account planned treatment after discharge; and (2) for the expected prognosis in the hypothetical situation had the patient not been admitted to hospital or treated elsewhere for the current health problem. Long-term QoL gain (LQG) was then calculated as the difference between these two assessments to find the improvement in QoL attributable to the hospital stay. The two assessments were made with the time trade-off instrument (TTO) which gives a measure of QoL in the interval from 0 (corresponding to coma or death) to 1 (corresponding to full health) [17]. When using the TTO, the experts first estimated the patient's remaining lifetime. They then decided how much of this they would have been willing to exchange for perfect health up to the time of death, had they been in the patient's situation. The lifetime left after this trade-off divided by the lifetime before trade-off is the TTO assessment of the patient's mean QoL. Details of the procedure have been given previously [12] and an example of its use can be found in Appendix A.

Short-term quality of life gain. Because the patient's remaining lifetime was used as the starting point for trade-off, the TTO's sensitivity for improvement in QoL of short duration relative to the remaining lifetime was limited. To compensate for this, the experts

also classified the improvement in QoL during the hospital stay or shortly after discharge in the categories no, low, intermediate or high gain relative to the expected QoL without admission. As some patients may have experienced short-term QoL gains which were unspecific effects of hospitalization, a further distinction was made between health-related short-term QoL gain (HSQG) and non-health-related short-term QoL gain (NHSQG). The former was defined as QoL gain resulting from any specific medical intervention or care, e.g. the relief of mental or somatic symptoms such as pain, nausea or depression. NHSQG was defined as all other types of QoL gain, e.g. when the hospital stay provided shelter for a homeless person or relief from a difficult social situation.

Evaluation protocol

The admissions were first assessed by each expert individually. Agreement between the three members of each panel was defined to exist when (1) the difference between the maximum and minimum LQG estimates was 0.20 or less, and (2) the HSQG and NHSQG assessments did not differ by more than one category. Otherwise, the estimates were discussed in a meeting and revised. Their median was taken to represent the panel's assessment.

Statistical methods

Ninety-five per cent confidence intervals (CI) of statistical parameters were estimated with the bootstrap algorithm [18]. The bootstrap distribution was obtained with a Monte Carlo simulation by drawing 10 000 random resamples of size 422 with replacement from the original observations.

A high proportion of observations had the value zero for the dependent variable LQG. In the multivariate linear regression analysis, the variance of the residuals was therefore not constant. Since this problem cannot be solved by transforming the dependent variable, the bootstrap algorithm was chosen for estimating confidence intervals for the regression coefficients as well. Their bootstrap distributions were found by calculating the least-squares estimates of the coefficients for each of the 10 000 resamples [18]. All confidence intervals were estimated with the bias-corrected and accelerated method described by Efron and Tibshirani [18].

The weighted kappa statistic was used for assessing agreement between the expert panels [19]. The squares of the number of categories of disagreement were used as weights when calculating the statistic.

This study was approved by the Regional Ethics Committee and the Norwegian Data Inspectorate.

Results

Agreement between the expert panels

For health-related short-term QoL gain (HSQG), the weighted kappa statistic for agreement between the two expert panels of admissions in group 1 was 0.70 (95% CI = 0.62–0.79; n = 57). For non-health-related short-term QoL gain (NHSQG), it was 0.08 (95% CI from −0.20 to +0.35; n = 57). Thus, the agreement for the first measure was good, whereas the second was no better than expected from chance [20].

The results of the investigation of agreement for long-term QoL gain (LQG) have been published previously. The mean difference between the panels' assessments was 0.02 (95% CI = 0.00–0.04; n = 57). When classified in categories of no, low and high gain, the weighted kappa statistic was 0.63 (95% CI = 0.45–0.80; n = 57), i.e. good agreement [12].

Assessments of groups 2 and 3 (n = 422)

These groups were assessed by only one of the two expert panels. For the rest of the analyses, they were pooled, giving n = 422. Of these, 160 (38%) were women, and 262 (62%) men. The mean age was 61.6 years, for women 61.0 years (range 16–94) and for men 61.9 years (range 15–90). A total of 152 (36%) were elective admissions, and 270 (64%) were emergency admissions: 20 (4.7%) patients died in the hospital. The mean length of stay was 8.6 days (SD = 20.5).

ICD9-codes were truncated to three digits. Related diagnoses were merged so that each diagnostic group included 10 admissions or more (Table 1).

Long-term quality of life gain. The mean LQG for all admissions was 0.06 (95% CI = 0.05–0.07) (Table 1). The distribution of LQG is shown in Fig. 1: 247 (59%) patients had LQG ≤ 0.00 (n = 422). Three of these had negative LQG. A 66-year-old man, who was admitted in a coma with cerebral haemorrhage, and who was discharged to a nursing home

Table 1 Mean long-term quality of life gain (*LQG*) estimated with the time trade-off method, and median health-related short-term quality of life gain (HSQG) from admissions to a department of internal medicine ($n = 422$)

	ICD9-code	Number of patients (%)	Mean LQG (95% confidence interval*)	Median HSQG
Total		422 (100%)	0.06 (0.05–0.07)	Low
Sex				
Men		262 (62%)	0.06 (0.05–0.08)	Low
Women		160 (38%)	0.06 (0.04–0.08)	Low
Age group				
<50 years		93 (22%)	0.04 (0.02–0.06)	Low
50–69 years		180 (43%)	0.05 (0.04–0.07)	Low
≥70 years		149 (35%)	0.09 (0.06–0.11)	Low
Admission category				
Elective		152 (36%)	0.04 (0.03–0.04)	Low
Emergency		270 (64%)	0.08 (0.06–0.10)	Moderate
Diagnostic group				
Infectious diseases	001–139	17 (4%)	0.11 (0.02–0.20)	Moderate
Malignant diseases	140–208	42 (10%)	0.06 (0.03–0.08)	Low
Endocrinological diseases	240–259	11 (3%)	0.20 (0.06–0.37)	Moderate
Acute myocardial infarction	410	30 (7%)	0.06 (0.01–0.13)	Low
Angina pectoris	411–414	85 (20%)	0.03 (0.02–0.04)	Low
Other heart diseases	420–429	45 (11%)	0.11 (0.07–0.16)	Low
Cerebrovascular diseases	430–438	21 (5%)	0.01 (−0.02–0.03)	Low
Pneumonia and influenza	480–487	16 (4%)	0.18 (0.08–0.26)	Moderate/high
Chronic obstructive pulmonary disease	496	20 (5%)	0.04 (0.01–0.10)	Moderate
Hepatobiliary/pancreatic diseases	570–579	13 (3%)	0.07 (0.01–0.14)	Low
Undiagnosed symptoms	780–789	30 (7%)	0.01 (0.00–0.01)	Low
Other		92 (22%)	0.06 (0.04–0.09)	Low

*Estimated with the bootstrap algorithm from 10 000 resamples.

whilst still in coma, had a LQG of −0.17. An 18-year-old man admitted for syncope and treated with a beta-blocker for a possible long QT-syndrome, although there was serious doubt about the diagnosis, scored −0.01. A 60-year-old man with angina pectoris in New York Heart Association class II, treated with percutaneous transluminal coronary angioplasty, scored −0.03. The procedure was successful, and the reason for the negative LQG was not obvious from the summary.

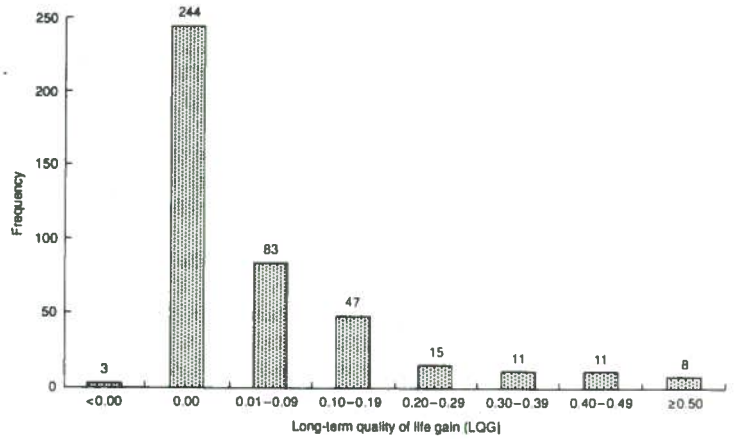


Fig. 1 Distribution of long-term quality of life gain (*LQG*) from admissions to a department of internal medicine as estimated with the time trade-off technique ($n = 422$).

Table 2 The eight admissions with long-term quality of life gain (LQG) greater than or equal to 0.50 ($n = 422$)

Sex	Age in years	ICD9 code	Admission category	LQG	Details
Female	77	250	Emergency	0.90	Patient with diabetes mellitus treated for hypoglycaemic coma
Male	74	410	Emergency	0.80	Pulmonary oedema caused by coronary heart disease treated with drugs and mechanical ventilation
Male	61	425	Emergency	0.65	Pulmonary oedmea with atrial fibrillation caused by cardiomyopathia treated with digitalis and other drugs
Female	67	428	Emergency	0.61	Incipient pulmonary oedema caused by coronary heart disease treated with drugs
Male	61	296	Emergency	0.53	Psychotic patient treated for dehydration and hypothermia, antidepressive medication initiated
Female	18	036	Emergency	0.50	Successfully treated meningococcal septicaemia
Male	63	038	Emergency	0.50	Patient with urosepsis treated with antibiotics
Female	86	711	Emergency	0.50	Infectious arthritis of the shoulder treated with antibiotics

The eight patients with LQG ≥ 0.50 accounted for 19% of the total LQG in the material (Table 2).

Short-term quality of life gain. The HSQG is shown in Table 1. The median for all admissions was low HSQG. Of the 247 admissions with LQG ≤ 0.00, two-thirds had some degree of HSQG (four high, 59 intermediate and 105 low).

A total of 79 (19%) had no health-related QoL gain at all, either LQG or HSQG ($n = 422$). Of these admissions, expert panel A judged 3% to have had intermediate, 28% low and 69% no NHSQG. The corresponding percentages for expert panel B were 3, 5 and 93%, respectively. Neither of the panels considered any of these 79 admissions to have had high NHSQG.

To summarize, 41% had LQG with or without HSQG, 40% had HSQG with or without NHSQG, and 19% had no gain or only NHSQG ($n = 422$) (Fig. 2).

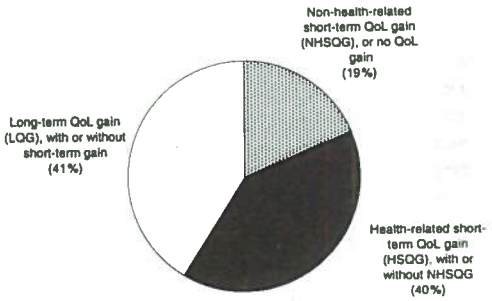


Fig. 2 Gain in quality of life from admissions to a department of internal medicine ($n = 422$).

Regression analysis

The effects on LQG of sex, age, admission category and diagnosis were investigated with a multivariate linear regression analysis (Table 3). Dummy variables were used for the diagnostic groups with 'other' as reference. 'Endocrinological diseases' and 'pneumonia and influenza' predisposed for higher gain, and 'undiagnosed symptoms' and 'cerebrovascular diseases' for lower gain than the reference. Emergency admissions and higher age were also significantly associated with higher gain.

Discussion

The most important result of this investigation was the uneven distribution of QoL gain, and in particular the very low gain for a high percentage of the admissions. Whilst a minority (2%) had gains of 0.50 or higher as measured with the time trade-off technique (Table 2), 19% of the admissions resulted in no health-related QoL gain (Fig. 2).

An attempt was made to find out whether these 19% had had QoL gain that was not health-related, but it failed because of lack of agreement between the experts for the NHSQG measure. It cannot be ruled out that these patients did achieve some improvement in quality of life, but neither expert panel estimated more than low NHSQG for more than 95% of them. Because this type of gain was defined as not resulting from specific medical interventions, it can be assumed that they could have achieved the same benefit without hospitalization.

Another 40% had health-related QoL gain of too

Table 3 Multivariate linear regression analysis of long-term quality of life gain (LQG) as estimated with the time trade-off technique ($n = 422$)

Independent variables*	Parameter estimate	95% confidence interval**
Intercept	−0.02	−0.07–0.02
Sex (0, male; 1, female)	−0.01	−0.03–0.02
Age	0.0011	0.0004–0.0020
Admission category (0, elective, 1, emergency)	0.03	0.01–0.05
Disease category		
Other	0.00	
Infectious diseases	0.06	−0.03–0.15
Malignant diseases	−0.01	−0.05–0.03
Endocrinological diseases	0.15	0.03–0.37
Acute myocardial infarction	−0.02	−0.06–0.07
Angina pectoris	−0.02	−0.05–0.00
Other heart diseases	0.04	−0.01–0.09
Cerebrovascular diseases	−0.08	−0.13 to −0.05
Pneumonia and influenza	0.10	0.03–0.21
Chronic obstructive pulmonary disease	−0.03	−0.08–0.02
Hepatobiliary/pancreatic diseases	0.02	−0.05–0.08
Undiagnosed symptoms	−0.06	−0.09 to −0.03

Adjusted R-square = 0.13.
*The disease category 'other' serves as a reference for the dummy variables of the disease categories.
**Estimated with the bootstrap algorithm from 10 000 resamples.

short a duration to be detected by the TTO technique. This group consisted of patients who had experienced relief from, for example, pain faster than they would have, had they not been treated. For most of them, the estimated gain was low. Even so, this kind of benefit is an important part of what has been called the Samaritan function of health care [14], and must continue to be an essential task of hospitals.

Forty-one per cent of the admissions resulted in gains detectable by the TTO method. The number of patients experiencing a certain QoL gain was inversely related to the amount of gain (Fig. 1). A few patients with gains of 0.50 or more had been successfully treated for life-threatening conditions with severe reductions in quality of life (Table 2).

Loss in quality of life

Only three patients (0.7%) had negative LQG even if our implementation of the TTO instrument allowed for both positive and negative gains. The frequency of adverse events in departments of general medicine in other studies has varied from 3.6 to 36%, which probably reflects differences in definitions and methods [21, 22]. Because the TTO instrument had limited sensitivity for positive gain, the same may have applied to negative gain, but it seems unlikely that the panels would have missed adverse events with major lasting negative effects on QoL.

We had no instrument for measuring short-term QoL losses, which therefore could have been experienced by some of the patients. Several kinds of treatments are known to reduce QoL temporarily to gain life expectancy or QoL in the long run, e. g. the treatment of malignant neoplasms with cytostatics.

Factors predisposing for gain in quality of life

From the perspective of priority setting, it is important to identify factors associated with high QoL gain. In the multivariate linear regression analysis, some of the diagnostic categories were significant regressors (Table 3). Because the groups were heterogeneous and did not take comorbidity into account, these results should be interpreted with caution. Even so, it is noteworthy that patients with symptoms without any specific diagnosis had significantly lower gain than the reference. The same applied to cerebrovascular diseases, for which effective forms of treatment in the acute phase are only now starting to emerge.

High age also predisposed for higher QoL gain. There has been a debate about how to contain the costs of the rising use of acute-care hospitals by old patients [23]. With regard to QoL, our results indicate that it is not correct to limit access to health care on the basis of high age alone. When considering

whether to admit a patient with the intention of improving QoL, high age should weigh in favour of admission rather than the opposite.

Since university hospitals often have a lower percentage of emergency admissions than other hospitals, the finding that these admissions were associated with higher gains might imply that the mean LQG could be higher in these hospitals. A higher mean age would also contribute to this tendency. However, because the diagnostic categories are important regressors for LQG, these effects could be counteracted by differences in case mix.

Limitations

The most important limitation of the present study was that its design did not allow patient self-assessment of QoL. This issue has been discussed in detail earlier [12]. It can hardly be denied that the patients' assessments must be the gold standard when it comes to measuring their own QoL. However, it is also clear that it is the doctors' assessments of how the patients experience different health states that ultimately determine which diagnostic and therapeutic interventions will be chosen. The good results of the agreement study indicate that there is consensus between doctors about QoL gain from hospital stays, except for NHSQG.

Another important limitation concerns the generalizability of our results. Although there is little reason to believe that the results would have been much different in other departments of internal medicine in the developed world, generalization to other parts of medical care is less straightforward. However, the diseases treated by internists include many of those with most severe prognoses. The potential for QoL gain for patients admitted to departments of internal medicine is therefore probably at least as great as for patients in other departments or in primary care.

Conclusions

Based on expert judgement, 81% of the admissions to a department of internal medicine resulted in some improvement in health-related quality of life. The gains were unevenly distributed. Half of these patients had only short-term improvement in their QoL, whilst a minority had high gains corresponding to the successful treatment of life-threatening conditions. The remaining 19% had either no improvement in QoL or improvement which had no direct relationship to specific medical interventions and which probably could have been achieved without hospital admission. Diagnosis was the most important determinant of gain, but high age and emergency admissions were also independently associated with higher gain.

Acknowledgements

This study was supported by grants from the Norwegian Medical Association Funds for Quality Improvement and the University Hospital of Tromsø Research Fund.

References

1 Wennberg J, Gittelsohn A. Small area variations in health care delivery. *Science* 1973; 182: 1102–8.
2 Roos NP, Roos LL. High and low surgical rates: risk factors for area residents. *Am J Public Health* 1981; 71: 591–600.
3 Andersen TF, Mooney G, eds. *The Challenges of Medical Practice Variations*. London: MacMillan Press, 1990.
4 Chassin MR, Brook RH, Park RE, Keesey J, Fink A, Kosecoff J *et al.* Variations in the use of medical and surgical services by the Medicare population. *N Engl J Med* 1986; 314: 285–90.
5 Chassin MR, Kosecoff J, Park RE, Winslow CM, Kahn KL, Merrick NJ *et al.* Does inappropriate use explain geographic variations in the use of health care services? A study of three procedures. *JAMA* 1987; 258: 2533–7.
6 Winickoff RN, Fischer MA, August BJ. Appropriateness of short-stay admissions for procedures in six Veterans Affairs hospitals. *QRB Qual Rev Bull* 1991; 17: 386–91.
7 Payne SM. Identifying and managing inappropriate hospital utilization: a policy synthesis. *Health Serv Res* 1987; 22: 709–69.
8 Siu AL, Sonnenberg FA, Manning WG, Goldberg GA, Bloomfield ES, Newhouse JP *et al.* Inappropriate use of hospitals in a randomized trial of health insurance plans. *N Engl J Med* 1986; 315: 1259–66.
9 McKeown T. *The Role of Medicine. Dream, Mirage or Nemesis?*, 2nd edn. Oxford: Basil Blackwell, 1979.
10 McKinlay JB, McKinlay SM, Beaglehole R. A review of the evidence concerning the impact of medical measures on recent mortality and morbidity in the United States. *Int J Health Serv* 1989; 19: 181–208.
11 Evans RG, Barer ML, Marmor TR, eds. *Why Are Some People Healthy and Others Not?* New York: Walter de Gruyter, 1994.
12 Eriksen BO, Almdahl SM, Hensrud A, Jæger S, Kristiansen IS, Mürer FA *et al.* Assessing health benefit from hospitalization: agreement between expert panels. *Int J Technol Assess Health Care* 1996; 12: 126–35.
13 Eriksen BO, Kristiansen IS, Nord E, Pape JF, Almdahl SM, Hensrud A *et al.* Does admission to a medical department improve patient life expectancy? *J Clin Epidemiol* 1997; 50: 987–95.
14 McDermott W. Medicine: the public good and one's own. *Perspect Biol Med* 1978; 21: 167–87.

15 Levine S, Feldman JJ, Elinson J. Does Medical Care Do Any Good? In: Mechanic D, ed. *Handbook of Health, Health Care and the Health Professions*. New York: Free Press, 1983; 394–404.
16 Anderson OW, Morrison EM. The worth of medical care: a critical review. *Med Care Rev* 1989; 46: 121–55.
17 Torrance GW. Measurement of health state utilities for economic appraisal. *J Health Economics* 1986; 5: 1–30.
18 Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. New York: Chapman & Hall, 1993.
19 Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull* 1968; 70: 213–20.
20 Fleiss JL. *Statistical Methods of Rates and Proportions*, 2nd edn. New York: John Wiley & Sons, 1981.
21 Brennan TA, Leape LL, Laird NM, Hebert L, Localio AR, Lawthers AG *et al.* Incidence of adverse events and negligence in hospitalized patients. *N Engl J Med* 1991; 324: 370–6.
22 Steel K, Gertman PM, Crescenzi C, Anderson J. Iatrogenic illness on a general medical service at a university hospital. *N Engl J Med* 1981; 304: 638–42.
23 Lamont CT, Sampson S, Matthias R, Kane R. The outcome of hospitalization for acute illness in the elderly. *J Am Geriatr Soc* 1983; 31: 282–8.

Received 17 November 1997; accepted 31 March 1998.

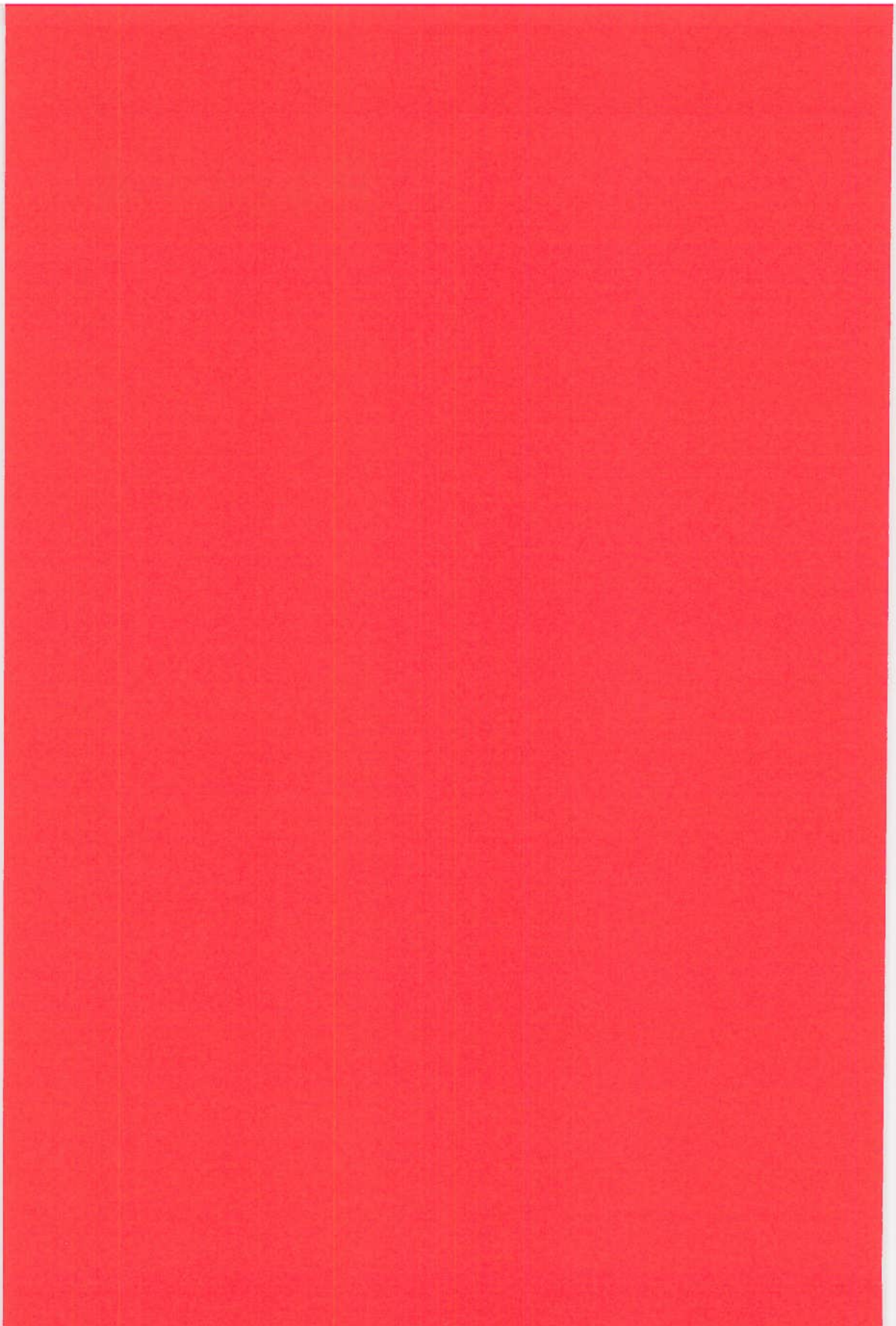
Correspondence: Bjørn Odvar Eriksen, Department of Medicine, University Hospital of Tromsø, 9038 Tromsø, Norway (fax: + 47 77626863; e-mail: medboe@rito.no).

Appendix: An example of assessment of long-term quality of life gain with the TTO method

A 63-year-old man was admitted for urosepsis. The expert panel estimated that he would have lived only 20 days in the hypothetical situation without hospital admission or treatment elsewhere. Considering his expected quality of life, they agreed that they would have been willing to give up half of these in exchange for living for only 10 days but in perfect health, i.e. to trade off lifetime in exchange for quality of life. In this situation, his mean QoL would have been the ratio between the lifetime after the trade-off and the lifetime before the trade-off, i.e. 0.50.

The expert panel then made the same assessment for the situation after he had been successfully treated in hospital. They expected a remaining lifetime of 12 years and a quality of life so good that they would not have been willing to trade off any lifetime to improve it. Thus, in this situation his QoL was 1.00. The gain in QoL attributable to the hospital stay is found by subtracting from this the value found without hospital admission, resulting in a long-term quality of life gain of 0.50.

Paper 4



THE COST OF INAPPROPRIATE ADMISSIONS: A STUDY OF HEALTH BENEFITS AND RESOURCE UTILISATION IN A DEPARTMENT OF INTERNAL MEDICINE

Running headline: The cost of inappropriate admissions

Bjørn O.Eriksen MD[1]
Ivar S. Kristiansen MD MPH[2]
Erik Nord PhD[3]
Jan Fr. Pape MD MPH[4]
Sven M. Almdahl MD[5]
Anne Hensrud MD MPH[6]
Steinar Jæger MD[7]

[1]Dept. of Medicine, University Hospital of Tromsø, Tromsø, Norway
[2]Institute of Community Medicine, University of Tromsø, Tromsø, Norway; present affiliation Centre for Health and Social Policy, Odense, Denmark
[3]National Institute of Public Health, Oslo, Norway
[4]Dept. of Medicine, University Hospital of Tromsø, Tromsø, Norway; present affiliation Dept. of Medicine, Telemark Central Hospital, Skien , Norway
[5]Dept. of Surgery, University Hospital of Tromsø, Tromsø, Norway; present affiliation Dept. of Surgery A, National Hospital, Oslo, Norway
[6]Kommunelegekontoret i Bardu, Bardu, Norway
[7]Dept. of Medicine, Nordland Central Hospital, Bodø, Norway

Abstract

Objectives: High rates of inappropriate hospital admissions have been found in numerous studies, suggesting that a high percentage of hospital resources are, in effect, wasted. The degree to which this is true depends on how costly inappropriate admissions are compared to other admissions. This study aimed to estimate both the percentage and cost of inappropriate admissions.

Setting: Department of internal medicine at a teaching hospital.

Subjects: Consecutively admitted patients during a six-week study period.

Main outcome measures: Assessments of inappropriateness were based on estimates of health benefit and necessary care level. These estimates were made by expert panels using a structured consensus method. Health benefit was estimated as gain in quality-adjusted life years, or degree of short-term improvement in quality of life during or shortly after the hospital stay. The direct costs to the hospital of each stay were estimated by allocating the costs of labor, "hotel" and overhead according to length of stay and adding to this the cost of ancillary resources used by each individual patient.

Results: 422 admissions were included. The 102 (24%) judged to be inappropriate had a lower mean cost (US$ 2,532) than the other 320 (US$ 5,800) (difference 3,268; 95% confidence interval 1,025 to 5,511). The inappropriate admissions accounted for 12% of the total costs.

Conclusions: Denying care for inappropriate admissions does not generate cost reductions of the same magnitude. Policy makers should be cautious in projecting the cost savings potential of excluding inappropriate admissions.

Keywords

hospital, health benefit, cost analysis, quality-adjusted life years, internal medicine, quality of life

Introduction

An "inappropriate hospital admission" can be defined as an admission that does not result in any significant benefit for the patient, or which results in benefit which could have been obtained on a lower care level. Studies from different countries have almost invariably found high rates of such admissions, with most reported percentages in the range of 10 to 25 [1-8]. One reason for the interest in inappropriate admissions has been the belief that they represent a potential for proportional cost reductions. However, this hypothesis depends on the assumption that inappropriate admissions are as costly as beneficial ones, which has so far not been investigated.

In the Tromsø Medical Department Health Benefit Study, health benefits as judged by expert panels have been studied in a department of internal medicine. We have previously reported that 61% of the patients admitted had no gain in life expectancy [9] and 19% no gain in quality of life [10]. In the present study, the benefits of these patients were recalculated in terms of quality-adjusted life years (QALY), and necessary care level assessed to estimate the percentage of inappropriate admissions. In addition, the costs of all stays were estimated. The primary aim of the study was to investigate the assumption that significant savings could have been obtained by denying care for inappropriate admissions. Second, we

wanted to examine the potential for savings by reducing the number of admissions with the lowest health benefits as well. This was done by estimating cost according to degree of health benefit.

Ideally, costs should have been estimated from the societal perspective, since many patients would have been treated elsewhere if they had not been admitted. However, for the heterogeneous group of patients admitted to departments of internal medicine, the alternatives to care in hospital are numerous. It was therefore not feasible to estimate costs for alternative care. Instead, an analysis was performed to explore the potential for cost reductions from the hospital's perspective.

Material and Methods

Subjects

During a six-week period from 1st February 1993, all admissions to the department of internal medicine at the University Hospital of Tromsø were eligible for inclusion in the study. Patients transferred from other university hospitals (n=3), patients admitted for evaluation or continuation of treatment started during a previous stay (n=27) and patients included in drug trials (n=2) were excluded, as well as one patient whose medical record could not be found. Nine planned readmissions were merged with the primary admission. A 10% random sample of the patients was used to study the extent of agreement between the two expert panels recruited for the study [11]. The remaining 422 admissions were used for the present investigation.

The two expert panels each consisted of an internist, a surgeon and a general practitioner. Each admission was randomly assigned to assessment of health benefit by one of them.

The study was approved by the Regional Ethics Committee and the Norwegian Data Inspectorate.

Health benefit and necessary care level

Health benefit assessments were made from summaries containing the patient's complete medical history and all data from the current stay. The time trade-off method [12] was used for estimating the gain in healthy-years equivalents (HYE) from the hospital stay relative to a hypothetical situation where the patient had not been admitted or treated elsewhere. HYE is a measure of life years adjusted for quality of life where 1 HYE represents one year in full health. The time trade-off method finds the number of HYE which the patient considers equivalent to living the rest of his life with reduced quality of life because of disease. The measure "healthy-years equivalents" is almost equivalent to "quality-adjusted life years" (QALY), and the term QALY will be used in this paper [13].

The time trade-off instrument has limited sensitivity for improvements in quality of life of short duration relative to the remaining lifetime, e.g. for the relief of symptoms associated with acute illness. To compensate, the experts also graded the improvement in health-related quality of life during the hospital stay or shortly after discharge in the categories no, low, intermediate or high gain. Details about the assessments of quality of life in this study have been published previously [10]. The expected outcome of planned treatment after discharge was taken into account when assessing health benefits.

The experts also assessed whether a patient with health benefit could have obtained the same benefit in primary care or in an outpatient clinic. If this was the case, or if an admission resulted in no health benefit, it was defined as inappropriate, otherwise as appropriate.

A structured consensus method was used for making the estimates [14]. The admissions were first assessed by each expert individually. When there was disagreement according to predefined criteria, the case was discussed in a meeting of the three members of the panel. After revision of the individual estimates, the median was taken to represent the panel's assessment. Further details of the method, a discussion of methodological problems and results about its reliability have been published previously [11].

Cost of hospital stays

Direct costs incurred during the stays in the department of internal medicine were estimated from the perspective of the hospital. For each patient, costs were estimated on the basis of unit costs and utilization of services.

Overhead costs were allocated to the service and clinical departments according to the step down allocation method [15]. The allocation basis used was the number of employees, square footage, number of admitted patients or number of patient-days as

appropriate. In the department of internal medicine, physician salaries were apportioned to wards and services according to the actual assignments of doctors in 1993.

The fee schedule for outpatients was used for calculating the unit costs for the service departments. For each department, the charges for the total production in 1993 was calculated as if all services had been paid according to this schedule. The total actual costs of the departments were then divided by these amounts to obtain cost-to-charge ratios which were multiplied by the outpatient fees to find the unit cost of specific services. For some services, outpatient clinic fees did not exist, and estimates of unit costs from an investigation in a similar hospital were used [16]. Utilization of diagnostic and therapeutic services for individual patients were recorded from computerized and manual databases (radiology, clinical chemistry, endoscopies, cardiologic interventions, hemodialysis, occupational therapy, blood components, etc.). In the following, the cost of these services and of pharmaceuticals will be termed "ancillary costs".

The costs of pharmaceuticals were set at the prices charged by the hospital pharmacy. Only drugs having a total cost of more than 1% of the department's total drug costs in 1993 were registered for the individual patient. The costs of other drugs were apportioned according to the length of stay for each ward separately, as were also nursing and "hotel" costs.

Capital costs of buildings and land are not included in the accounts of the hospital, and were excluded from the cost analysis. The cost of major equipment was accounted for directly without annual depreciation, which is not routinely calculated in the hospital accounts. Because these costs will vary from year to year, they were averaged over the years 1992, 1993 and 1994 for each department.

All costs were measured in 1993 NOK and converted according to the exchange rate US$1=NOK 7.50.

Statistical methods

Ninety-five per cent confidence intervals of statistical parameters were estimated with the bootstrap algorithm [17]. Multivariate linear regression analysis was performed with the SAS statistical package (SAS Institute, Cary, NC).

Sensitivity analysis

The difference between the mean costs of appropriate and inappropriate admissions was explored in a probabilistic sensitivity analysis where all unit costs were varied simultaneously by drawing them from logistic-normal probability density distributions in 10,000 runs of a Monte Carlo simulation according to the method described by Doubilet et al [18]. For each unit cost, the parameters of this distribution were calculated from the estimated unit cost, and from

lower and upper bounds of 2.5% and 197.5% of the estimated unit cost respectively.

Results

Inappropriate admissions and health benefits

One hundred-two (24%) of the 422 admissions were inappropriate, and 115 (27%) resulted in only short-term improvement of quality of life during or shortly after the stay. Two hundred-five (49%) had benefits measured as QALY (Table 1). The mean gain in QALY was 2.3 per admission.

Of the 115 admissions with gain in health-related short-term quality of life, 74 had low, 38 moderate and 3 high gain (Table 1). Clinical details of the six patients in the low gain category with the highest costs are listed in Table 2.

Two admissions resulted in a QALY loss, i.e. that the patient's health was made worse by the hospital stay (-0.1 and -0.6 respectively). These admissions were defined as inappropriate. Details of the admissions with health loss have been given previously [9,10].

Cost analysis

The total cost of the 422 admissions was US$ 2.1 million (Table 3). Overhead (32%) and nursing costs (27%) made up the largest proportions of this total. For both the appropriate and the inappropriate admissions, the ancillary costs were 29% of the total.

Table 4 shows the mean cost according to gender, age, admission type and diagnostic category. The mean cost of the inappropriate admissions (US$ 2,532) was lower than for the appropriate (US$ 5,800) (difference 3,268; 95% confidence interval 1,025 to 5,511). In a multivariate linear regression analysis of logarithmically transformed cost with appropriateness, gender, age, admission category and dummy variables for the diagnostic categories as independent variables, appropriate admissions were associated with higher cost (P<0.001) (Table 5). The diagnostic categories "angina pectoris" (P=0.013) and "undiagnosed symptoms" (P=0.028) were associated with lower costs. No interactions between appropriateness and the other variables were detected (P>0.05).

The relationship between appropriateness, health benefit and cost is further explored in Table 1. The 24% of inappropriate stays accounted for 12% of the total costs. The 42% of stays which were either inappropriate or had only low, health-related short-term quality of life gain, together accounted for 25% of the costs. The mean length of stay for inappropriate admissions was 4.3 days (95% confidence interval 3.1 to 5.8), for appropriate admissions 10.0 days (95% confidence interval 7.9 to 13.1).

Sensitivity analysis

When the unit costs were varied simultaneously in a Monte Carlo simulation of 10,000 runs, none resulted in a higher mean cost for inappropriate than for appropriate admissions.

Discussion

Few investigators of "inappropriate hospital admissions" provide a definition of this term. Those who do, base their definition on the concept "health benefit", or just "benefit" [19]. We are not aware of any study of inappropriate admissions which has included a definition of "health benefit", or a description of methods for measuring it. Instead, assessments of health benefit have relied on implicit clinical judgment, either directly or through validation of instruments by expert physicians [2,19,20]. In the Tromsø Medical Department Health Benefit Study, a set of explicit criteria designed to be sensitive to all gains in life expectancy and health-related quality of life was used in a two-round consensus process. Definitions, descriptions of methods and results form the application of the instrument to consecutive admissions have been reported in previous publications from the study [9,10]. The instrument has been found reliable for a random sample of the included admissions [11].

According to the final assessment of the two panels, 24% of the admissions were inappropriate. Previously, we have discussed the possibility that the experts had overlooked benefits for some of these admissions [10]. In particular, we were concerned that some of the patients might have experienced improvements in quality of life from having a tentative diagnosis confirmed or excluded, even if this did not lead to improvement in health. The expert panels both estimated

that less than 5% of the inappropriate admissions had achieved more than the lowest degree of this type of benefit. Therefore, it is unlikely that more than a few of the inappropriately admitted patients had had improvements in quality of life that made hospitalization necessary.

However, this result was based on assessments of the patients' quality of life by physicians, and it could be argued that the rate of inappropriate admissions would have been different if it had been based on the patients' own assessments. The justification for our approach was that both inappropriate admissions and the costs of hospital stays are the results of decisions made by clinicians. These decisions will be determined by the clinicians' assessments of the health benefits for patients resulting from various alternatives. Accordingly, these assessments are relevant measures in investigations aiming to study the relationship between health benefits and costs.

The percentage of inappropriate admissions found in this study was comparable to those found in other studies, and confirms that there is a potential for reducing the number of admissions without loss in health benefits. However, the finding that the cost of these admissions was less than 50% of that of the others challenges the hypothesis that this would lead to savings of the same magnitude. Even if the experts had been biased towards considering patients

who had undergone costly interventions as appropriate, this cannot explain the entire difference in mean cost between the appropriate and other admissions. Also, the costs of interventions were included in the ancillary costs, which only accounted for 29% of the total costs. This would limit the effect of this type of bias on the difference in mean cost. The most important determinant of cost was length of stay, which was considerably shorter for the inappropriate admissions. One reason for this might have been that these patients were discharged earlier because their low potential for benefit was recognized soon after admission.

An attempt was made to identify subgroups of inappropriate admissions with especially high costs by testing for interactions between appropriateness and other variables in a multivariate regression analysis. However, although some groups had lower costs independently of appropriateness, we were not able to identify any group for which a reduction of inappropriate admissions would lead to a greater cost reduction than for others. The variables investigated were gender, age, admission category and diagnostic category, which specify a rather crude model relative to the detailed clinical information available about each patient. The result of the analysis does not exclude the possibility that a higher percentage of savings could be obtained by targeting more carefully defined subgroups of inappropriate admissions.

The 24% inappropriate admissions accounted for 12% of the costs. It should be noted, however, that we have estimated average costs of care. When estimating cost savings from admitting fewer patients, marginal costs, i.e. the additional cost of treating one more patient, are more relevant. Most of the labor costs, which represented about two thirds of the hospital's total costs in 1993, are fixed in the short run. Accordingly, the savings from excluding inappropriate admissions would have been much less than 12% in this time perspective. In the long run, all costs are variable, and the cost savings would have been in the order of 12%.

An important limitation of this study was that the cost analysis was made from the hospital's perspective and included only costs incurred during the hospital stays. Some of the patients denied hospital care would have been treated on a lower care level and incurred costs here. Consequently, the savings from excluding inappropriate admissions could have been lower from a societal than from the hospital's perspective. A cost analysis from the societal perspective would have been preferable, but estimating costs outside the hospital was not feasible in this study because it was difficult to make assumptions about alternative care for this heterogeneous group of patients. Coast et al estimated the potential for societal cost savings from alternative care for inappropriately admitted patients by assuming average speciality costs and the same duration of care as in the hospital [21]. It was concluded that

References

1. Payne SM. Identifying and managing inappropriate hospital utilization: a policy synthesis. *Health Serv Res* 1987; **22:** 709-769.

2. Siu AL, Sonnenberg FA, Manning WG, Goldberg GA, Bloomfield ES, Newhouse JP, et al. Inappropriate use of hospitals in a randomized trial of health insurance plans. *N Engl J Med* 1986; **315:** 1259-1266.

3. Coast J, Inglis A, Frankel S. Alternatives to hospital care: what are they and who should decide? *BMJ* 1996; **312:** 162-166.

4. Davido A, Nicoulet I, Levy A, Lang T. Appropriateness of admission in an emergency department: reliability of assessment and causes of failure. *Qual Assur Health Care* 1991; **3:** 227-234.

5. Fellin G, Apolone G, Tampieri A, Bevilacqua L, Meregalli G, Minella C, et al. Appropriateness of hospital use: an overview of Italian studies. *Int J Qual Health Care* 1995; **7:** 219-225.

6. Alonso J, Munoz A, Anto JM. Using length of stay and inactive days in the hospital to assess appropriateness of utilisation in Barcelona, Spain. *J Epidemiol Community Health* 1996; **50:** 196-201.

7. Bentes M, Gonsalves ML, Santos M, Pina E. Design and development of a utilization review program in Portugal. *Int J Qual Health Care* 1995; **7:** 201-212.

8. Perneger TV, Chopard P, Sarasin FP, Gaspoz JM, Lovis C, Unger PF, et al. Risk factors for a medically inappropriate admission to a Department of Internal Medicine. *Arch Intern Med* 1997; **157:** 1495-1500.

9. Eriksen BO, Kristiansen IS, Nord E, Pape JF, Almdahl SM, Hensrud A, et al. Does admission to a medical department improve patient life expectancy?. *J Clin Epidemiol* 1997; **50:** 987-995.

10. Eriksen BO, Kristiansen IS, Nord E, Pape JF, Almdahl SM, Hensrud A, et al. Does admission to a department of internal medicine improve patient quality of life?. *J Intern Med* 1998; **244:**397-404.

11. Eriksen BO, Almdahl SM, Hensrud A, Jæger S, Kristiansen IS, Mürer FA, et al. Assessing health benefit from hospitalization: Agreement between expert panels. *Int J Technol Assess Health Care* 1996; **12:** 126-135.

12. Torrance GW. Measurement of health state utilities for economic appraisal. *J Health Economics* 1986; **5:** 1-30.

13. Torrance GW, Feeny D. Utilities and quality-adjusted life years. *Int J Technol Assess Health Care* 1989; **5:** 559-575.

14. Jones J, Hunter D. Consensus methods for medical and health services research. *BMJ* 1995; **311:** 376-380.

15. Drummond MF, O'Brien B, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes.* Second ed. Oxford: Oxford Medical Publications; 1997.

16. Slåttebrekk OV. *Hva koster pasienten?* Trondheim: Norsk institutt for sykehusforskning; 1990. NIS-Rapport 1/90.

17. Efron B, Tibshirani RJ. *An Introduction to the Bootstrap.* New York: Chapman & Hall; 1993.

18. Doubilet P, Begg CB, Weinstein MC, Braun P, McNeil BJ. Probabilistic sensitivity analysis using Monte Carlo simulation. A practical approach. *Med Decis Making* 1985; **5:** 157-177.

19. Gertman PM, Restuccia JD. The Appropriateness Evaluation Protocol: A Technique for Assessing Unnecessary Days of Hospital Care. *Med Care* 1981; **19:** 855-871.

20. Strumwasser I, Paranjpe NV, Ronis DL, Share D, Sell LJ. Reliability and validity of utilization review criteria. Appropriateness Evaluation Protocol, Standardized Medreview Instrument, and Intensity-Severity-Discharge criteria. *Med Care* 1990; **28:** 95-111.

21. Coast J, Inglis A, Morgan K, Gray S, Kammerling M, Frankel S. The hospital admissions study in England: are there alternatives to emergency hospital admission? *J Epidemiol Community Health* 1995; **49:** 194-199.

22. Coast J. Appropriateness versus efficiency: the economics of utilisation review. *Health Policy* 1996; **36:** 69-81.

Correspondence address

Bjørn O. Eriksen MD

Dept. of Medicine, University Hospital of Tromsø, 9038 Tromsø, Norway

T. +47 776 26856 Fax: +47 776 26863 E-mail: medboe@rito.no

Alternative proof reader

Ivar S. Kristiansen MD

Bredo Stabells vei 4, 0853 Oslo, Norway

T. +47 22 95 02 72 Fax: +47 22 95 21 59 E-mail: ivarsk@online.no

Acknowledgements

This study was supported by grants from the Norwegian Medical Association Funds for Quality Improvement and the University Hospital of Tromsø Research Fund. We are grateful to F. A. Mürer, R. Robertsen and G. Thorsen for participating in the expert panels, to O. H. Førde for assistance in preparing this manuscript, and to G. Mooney and K. Rasmussen for valuable comments. We are also indebted to O. V. Slåttebrekk for reviewing the cost analysis, and to the following for extracting data from various hospital databases: O. B. Gamst, H. Jentoft, A. Lockert, H. M. Pettersen, T. Simonsen, L. Småbrekke, J. Størmer, R. Sørvoll, I. Tranung.

Table 1 Costs (US$), appropriateness and health benefit from admissions to a department of internal medicine (n=422)

		No (%)	Mean cost per patient (95% CI)[1]	Total costs (%)
Inappropriate admissions[2]		102 (24)	2,532 (1,963 to 3,225)	258,305 (12)
Gain in short-term health-related quality of life only[3]				
	Low	74 (18)	3,744 (2,363 to 5,834)	277,076 (13)
	Moderate	38 (9)	4,232 (3,060 to 6,367)	160,803 (8)
	High	3 (1)	6,731 (4,380 to 10,694)	20,193 (1)
Gain in quality-adjusted life years (QALY)				
	0-0.5 QALY	29 (7)	5,701 (4,219 to 7,972)	165,315 (8)
	0.5-1 QALY	17 (4)	7,209 (3,483 to 12,626)	122,551 (6)
	1-5 QALY	107 (25)	6,889 (4,873 to 9,439)	737,170 (35)
	5-10 QALY	31 (7)	7,038 (5,005 to 9,047)	218,182 (10)
	>=10 QALY	21 (5)	7,373 (4,313 to 10,657)	154,830 (7)
Total		422 (100)	5,010 (4,275 to 6,048)	2,114,425 (100)

[1] Bias-corrected and accelerated 95% confidence intervals calculated with the bootstrap algorithm from 1,000 runs of a Monte Carlo simulation

[2] Patients with no health benefit, or benefit that could have been obtained on a lower care level

[3] Patients with improvement in health-related quality of life during or shortly after the hospital stay too small to be registered as gain in QALY

Table 2 The six patients with the highest costs in the "low short-term quality of life gain" category

Sex	Age in years	ICD9 code	Admission category	Total cost in US$	Clinical details
Female	75	310	Elective	86,696	Multiple myeloma with skeletal affection. Dementia. Radiation treatment of pathological fracture. Stayed in hospital for more than 6 months because of insufficient nursing home capacity.
Female	62	434	Emergency	16,790	Acute cerebral infarction with left hemiparalysis. The patient did not want rehabilitation and was discharged to a nursing home. Antibiotic for urinary tract infection. Antidepressivum for depressive neurosis.
Male	52	515	Emergency	9,120	Died of diffuse alveolointerstitial pulmonary diease. Palliative treatment.
Male	66	434	Elective	8,680	Admitted for rehabilitation after cerebral infarction. Unsuccesful, partly because of depression.
Male	64	496	Emergency	8,124	Chronic obstructive pulmonary disease, admitted for acute exacerbation.Some improvment after drug treatment.
Male	72	431	Emergency	6,941	Hemorrhage in capsula externa dxt. with moderate paresis. Quick restoration of function.

Table 3 Total costs (US$) and patient-days for 422 admissions to a department of internal medicine according to appropriateness of admission

	Inappropriate admissions (%) (n=102)	Appropriate admissions (%) (n=320)	All admissions (%) (n=422)
Ancillary costs	74,202 (12)	534,514 (88)	608,716 (100)
Nursing labor cost	66,083 (12)	505,084 (88)	571,167 (100)
Physician labor cost	17,493 (13)	118,258 (87)	135,750 (100)
Overhead	87,404 (13)	593,472 (87)	680,876 (100)
"Hotel" costs	13,123 (11)	104,793 (89)	117,916 (100)
Total costs	258,305 (12)	1,856,120 (88)	2,114,425 (100)
Total patient-days	433 (12)	3191 (88)	3,624 (100)

Table 4 Costs (US$) from admissions to a department of internal medicine (n=422) according to appropriateness of admission

	ICD9-code	Inappropriate admissions[1]			Appropriate admissions			All admissions		
		No (%)	Cost of stay		No (%)	Cost of stay		No (%)	Cost of stay	
			Mean	95% CI[2]		Mean	95% CI[2]		Mean	95% CI[2]
Total		102 (24)	2,532	1,963 to 3,225	320 (76)	5,800	4,857 to 7,121	422 (100)	5,010	4,275 to 6,048
Sex:										
Men		63 (24)	2,889	2,153 to 3,881	199 (76)	5,290	4,375 to 6,405	262 (100)	4,713	3,953 to 5,642
Women		39 (24)	1,956	1,300 to 3,157	121 (76)	6,639	4,765 to 9,437	160 (100)	5,498	4,076 to 7,812
Age group:										
<50 years		28 (30)	1,616	1,247 to 2,220	65 (70)	4,973	3,622 to 6,603	93 (100)	3,962	2,957 to 5,138
50-69 years		37 (21)	2,709	1,829 to 4,078	143 (79)	4,685	3,881 to 5,711	180 (100)	4,279	3,611 to 5,117
>=70 years		37 (25)	3,049	1,949 to 4,607	112 (75)	7,705	5,626 to 11,467	149 (100)	6,549	4,903 to 9,434
Admission category:										
Elective		52 (34)	2,207	1,739 to 2,798	100 (66)	5,553	4,107 to 7,872	152 (100)	4,408	3,375 to 5,970
Emergency		50 (19)	2,871	1,842 to 4,176	220 (81)	5,913	4,838 to 7,713	270 (100)	5,349	4,444 to 6,880
Diagnostic category:										
Infectious disease	001-139	4 (24)	1,804	341 to 2,721	13 (76)	8,693	4,208 to 13,460	17 (100)	7,072	3,610 to 11,454
Malignant disease	140-208	10 (24)	6,130	2,972 to 10,002	32 (76)	6,147	3,836 to 9,118	42 (100)	6,143	4,271 to 8,433
Endocrinological disease	240-259	3 (27)	3,072	2,633 to 3,686	8 (73)	4,920	1,866 to 11,705	11 (100)	4,416	2,149 to 9,776
Acute myocardial infarction	410	6 (20)	1,958	495 to 3,554	24 (80)	4,970	3,926 to 6,418	30 (100)	4,368	3,425 to 5,633
Angina pectoris	411-414	18 (21)	1,892	1,224 to 2,873	67 (79)	3,229	2,698 to 3,773	85 (100)	2,946	2,499 to 3,415
Other heart disease	420-429	4 (9)	1,736	96 to 3,414	41 (91)	5,920	3,563 to 10,915	45 (100)	5,548	3,331 to 10,114
Cerebrovascular disease	430-438	4 (19)	4,255	1,068 to 12,474	17 (81)	11,577	6,434 to 17,836	21 (100)	10,183	6,079 to 15,693
Pneumonia and influenza	480-487	4 (25)	1,596	566 to 3,054	12 (75)	6,401	4,468 to 8,430	16 (100)	5,200	3,404 to 7,136
Chronic obstr. pulm. disease	496	4 (20)	1,704	415 to 4,541	16 (80)	3,467	2,220 to 5,375	20 (100)	3,115	2,084 to 4,691
Heapatobiliary/pancreatic disease	570-579	6 (46)	1,085	600 to 1,466	7 (54)	6,200	2,768 to 12,845	13 (100)	3,839	1,775 to 8,280
Undiagnosed symptoms	780-789	16 (53)	2,405	1,184 to 5,075	14 (47)	2,362	1,284 to 4,019	30 (100)	2,385	1,436 to 3,918
Other		23 (25)	2,288	1,460 to 3,620	69 (75)	7,581	4,723 to 12,388	92 (100)	6,258	4,124 to 9,756

[1] Patients with no health benefit, or benefit that could have been obtained on a lower care level

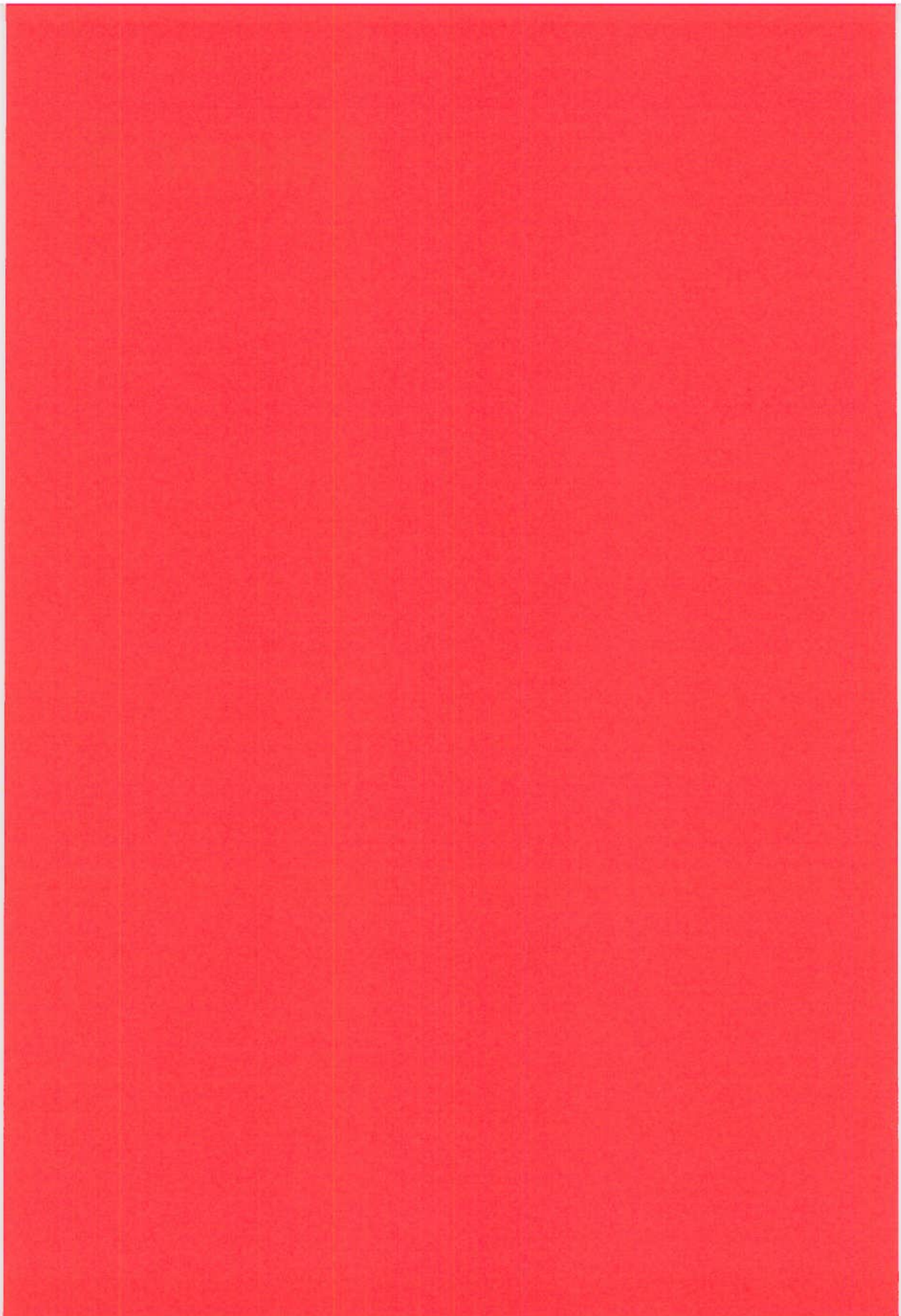
[2] Bias-corrected and accelerated 95% confidence intervals calculated with the bootstrap algorithm from 1,000 runs of a Monte Carlo simulation

able 5 Multivariate linear regression analysis of cost (logarithmically transformed) admissions to a department of internal medicine (n=422)

dependent variable		Parameter estimate	95% confidence interval	P-value
tercept		3.12	2.89 to 3.35	<0.001
ppropriateness (0=inappropriate, 1=appropriate)		0.35	0.23 to 0.47	<0.001
ex (0=M, 1=F)		-0.05	-0.16 to 0.05	0.307
ge		0.00	0.00 to 0.01	0.104
dmission category (0=elective, 1=emergency)		-0.10	-0.22 to 0.01	0.067
iagnostic category				
	Infectious diseases	0.19	-0.07 to 0.45	0.159
	Malignant diseases	0.01	-0.17 to 0.20	0.901
	Endocrinological diseases	0.04	-0.27 to 0.35	0.806
	Acute myocardial infarction	0.04	-0.17 to 0.25	0.696
	Angina pectoris	-0.20	-0.35 to -0.04	0.013
	Other heart diseases	-0.08	-0.27 to 0.10	0.378
	Cerebrovascular diseases	0.20	-0.04 to 0.44	0.100
	Pneumonia and influenza	0.13	-0.15 to 0.40	0.361
	Chronic obstr. pulm. disease	-0.12	-0.36 to 0.13	0.348
	Heapatobiliary/pancreatic diseases	-0.09	-0.38 to 0.20	0.554
	Undiagnosed symptoms	-0.23	-0.44 to -0.03	0.028
	Other	0.00	Reference	

djusted R^2 0.11; F=4.508; df=15, 406; P<0.001 (dummy variables used for the diagnostic categories)

Paper 5



COST SAVINGS AND HEALTH LOSSES FROM REDUCING INAPPROPRIATE ADMISSIONS TO A DEPARTMENT OF INTERNAL MEDICINE

Bjørn O.Eriksen MD[1]
Olav H. Førde MD[2]
Ivar S. Kristiansen MD MPH[3]
Erik Nord PhD[4]
Jan Fr. Pape MD MPH[5]
Sven M. Almdahl MD[6]
Anne Hensrud MD MPH[7]
Steinar Jæger MD[8]
Fred A. Mürer MD[9]

Correspondence and reprint requests: Bjørn O. Eriksen, Dept. of Medicine, University Hospital of Tromsø, 9038 Tromsø, Norway

[1]Dept. of Medicine, University Hospital of Tromsø, Tromsø, Norway
[2]Institute of Community Medicine, University of Tromsø, Tromsø
[3]Institute of Community Medicine, University of Tromsø, Tromsø, Norway; present affiliation Centre for Health and Social Policy, Odense, Denmark
[4]National Institute of Public Health, Oslo, Norway
[5]Dept. of Medicine, University Hospital of Tromsø, Tromsø, Norway; present affiliation Dept. of Medicine, Telemark Central Hospital, Skien , Norway
[6]Dept. of Surgery, University Hospital of Tromsø, Tromsø, Norway; present affiliation Dept. of Surgery A, National Hospital, Oslo, Norway
[7]Kommunelegekontoret i Bardu, Bardu, Norway
[8]Dept. of Medicine, Nordland Central Hospital, Bodø, Norway
[9]Dept. of Medicine, Rana Hospital, Mo, Norway

Summary

Background

The high rates of inappropriate hospital admissions found in many studies are commonly believed to represent a potential for significant cost reductions. However, this presumes that these patients can be identified before the hospital stay. The aims of this study were to investigate to what extent this is possible in a department of internal medicine, and to estimate the costs saved if patients judged at the time of admission to be inappropriately admitted, are denied care.

Methods

Consecutive admissions during a six week period were randomised for assessment by one of two expert panels. On the basis of the information available at the time of admission, the panels predicted the health benefit from the stays and the lowest necessary care level using a structured consensus method. Admissions with no benefit or with a lower necessary care level were defined as inappropriate. For each admission, a final judgement of appropriateness was made after discharge by the other panel which had access to all information collected during the stay. The predictions were then compared with these assessments as the gold standard. The direct costs to the hospital incurred during each stay were estimated.

Findings

The panels correctly classified 88% of the appropriate (n=320) and 27% of the inappropriate admissions (n=102). If the elective admissions predicted to be inappropriate had been excluded, 9% of the costs would have been saved and 5% of the gain in quality-adjusted life years (QALY) lost. The corresponding results for emergency admissions were 14% and 18%.

Interpretation

The savings obtained by excluding admissions predicted to be inappropriate were small relative to the health losses. High rates of such admissions do not necessarily imply that costs can be saved. Programs for reducing inappropriate health care should not be implemented without investigating their effects on both health outcomes and costs.

Keywords

cost effectiveness, hospital, health benefit, sensitivity and specificity, cost analysis, rationing

Introduction

Increasing health care costs have given rise to a variety of strategies for cost containment. One of them is to deny care when health benefits are negligible. It is commonly believed that the reduction of unnecessary or inappropriate health care would result in substantial savings [1-3]. In particular, this applies to inappropriate hospital admissions, for which high rates have been found in many countries [4-9]. However, the finding of a high rate of inappropriate admissions retrospectively does not necessarily indicate a potential for cost reductions. To reduce the number of such admissions and to obtain savings, clinicians must be able to identify them as inappropriate before or at the time of admission, that is, before diagnostic and therapeutic interventions are undertaken. The assumption that this is possible has, to our knowledge, not been investigated [10-12].

In the Tromsø Medical Department Health Benefit Study, 24% of the admissions to the department of internal medicine at a teaching hospital were found inappropriate [13]. This estimate was made by expert panels using a structured consensus method which has been found reliable for a random sample of the included patients [14]. In the present study, we investigated whether cost reductions could have been obtained by letting the expert panels predict appropriateness solely on the basis of information available at the time of admission. The aim of the study was twofold. First, to estimate the sensitivity and specificity of the these predictions. Second, to estimate the costs saved if they had been used for reducing the number of admissions and the potential health losses for patients falsely predicted not to need hospitalisation.

Methods

Subjects

In 1993, 5151 patients were admitted to the department of internal medicine at the University Hospital of Tromsø. During a six week period from 1st February 1993, all admissions were eligible for inclusion in the study. Patients transferred from other university hospitals (n=3), admitted for evaluation or continuation of treatment initiated during a previous stay (n=27) or admitted for inclusion in drug trials (n=2) were excluded, as well as one patient whose medical record could not be found. Nine planned readmissions were merged with the primary admission, resulting in 479 included admissions.

These admissions were randomly assigned to three groups with probabilities of 0.10 (group 1), 0.45 (group 2) and 0.45 (group 3). Two expert panels (A and B) were recruited, each consisting of an internist, a surgeon and a general practitioner who were all board-certified. For each admission in group 2 and 3, appropriateness was predicted at admission by one of the panels, and a final judgement of appropriateness made by the other panel after discharge (Figure 1). The admissions in group 1 were assessed by both panels after discharge to study inter-panel agreement [14].

The study was approved by the Regional Ethics Committee and the Norwegian Data Inspectorate.

Predictions of appropriateness

An admission was defined as appropriate if it resulted in health benefit which could not have been obtained on a lower care level. For prediction of health benefit, the experts were provided with the patient's complete medical history and the results of the physical examination as obtained at admission. No information about the course of the hospital stay after the time of admission was given. Using a method which has been described in more detail previously, the experts then made predictions of the health gain from the hospital stays in terms of healthy-years equivalents (HYE) [14]. HYE is a measure of life years adjusted for quality of life where 1 HYE represents one year in full health [15,16]. Although there are some theoretical differences between HYE and the more well-known "quality-adjusted life years" (QALY), the latter term will be used in this paper [17,18].

The measurement of gain in QALY has limited sensitivity for improvement in quality of life of short duration relative to the remaining lifetime. To compensate, the experts also predicted the improvement in health-related short-term quality of life during the hospital stay or shortly after discharge relative to the expected quality of life without admission [19]. Finally, they predicted whether patients with health benefits could have obtained the same benefit in primary care or in an outpatient clinic.

The predictions were first made by each expert individually, and then discussed in a meeting of the three members of each panel when there was disagreement according to predefined criteria. Further details of the method, a discussion of methodological problems and results regarding its reliability have been published previously [14].

For each admission, final assessments of health benefit and care level were made by the other panel after discharge. The results of these assessments, which in the following will be termed the observations, have been reported in detail previously [13,19,20]. The predicted and observed appropriateness of the admissions were determined from the assessments of health benefit and necessary care level. To estimate the sensitivity and specificity of the prediction that an admission would be appropriate, the predictions were compared with the observations as the gold standard. The formulae used were

$$\text{sensitivity} = \frac{\text{no. of admissions both predicted and observed to be appropriate} \times 100}{\text{no. of admissions observed to be appropriate}}$$

and

$$\text{specificity} = \frac{\text{no. of admissions both predicted and observed to be inappropriate} \times 100}{\text{no. of admissions observed to be inappropriate}}$$

Group 2 and 3 were pooled for this analysis.

Cost analysis

Direct costs in 1993 NOK (US$ 1= NOK 7.50) incurred by the patients during their stays in the department were estimated from the perspective of the hospital. Overhead costs were allocated to the service and clinical departments according to the step down allocation method [21]. For each patient, costs were estimated on the basis of unit costs and utilisation of services. Unit costs were estimated for the output of all service departments (radiology, microbiology, etc.). Utilisation of services was registered from hospital databases and the medical record for each individual patient.

The costs of nurse and physician labour and "hotel costs" were apportioned according to length of stay for each ward separately.

Further details of the cost analysis have been given previously [13].

Statistical methods

Multivariate logistic regression analyses were performed with the SAS statistical package (SAS Institute, Cary, NC).

95% confidence intervals of statistical parameters were estimated with the bootstrap algorithm, except for the logistic regression [22]. The kappa statistic was used for assessing agreement between the expert panels for assessments about the admissions in group 1 [23].

Results

Agreement between the expert panels

Group 1 (n=57), in which all patients were assessed by both expert panels after discharge, was used for estimating the agreement for judging that an admission was appropriate. The overall agreement was 0.75, the kappa statistic 0.41 (95% CI 0.15 - 0.68), *i.e.* fair agreement [24].

Prediction of appropriateness

Of the admissions in group 2 and 3 (n=422), the expert panels predicted that 66 (16%) would be inappropriate and 356 (84%) appropriate. The relationship between these predictions and the observations made by the panels after discharge is shown in Table 1. The panels were able to identify 88% of the appropriate but only 27% of the inappropriate admissions. In other words, the sensitivity and specificity of the prediction that an admission would be appropriate were 88% and 27% respectively. The sensitivities and specificities for subgroups are shown in Table 2.

To explore whether there was an association between the predictions and the observations, a logistic regression analysis was performed with the predictions as the dependent variable. The observations of appropriateness, gender, age and dummy variables for diagnostic categories were included as independent variables. Elective and emergency admissions were analysed separately. For emergency admissions, the fit of the model was poor (chi-square for covariates 17.56, d.f. 13, P=0.18). For elective admissions, the fit was better (chi-square for covariates 25.15, d.f. 10, P=0.005) (Table 3). Only the observation of appropriateness made after discharge and gender were significant regressors. No interaction between these two variables

was observed (P=0.22). Because the odds ratio for the observations is indicative of the panels' predictive abilities in this model, the absence of this interaction means that these abilities were the same for men and women. However, because of the gender variable, the sensitivity and specificity for the two sexes were different. Based on the crude data, the sensitivity for elective admission of men was 96% and of women 75%. The specificities were 21% and 50% respectively. Sensitivities and specificities estimated from the model were similar.

Clinical details of the 5 patients with the greatest predicted health benefits who were judged to be inappropriate after discharge, and of the 5 patients predicted to be inappropriate with the greatest health benefits, can be found in Table 4.

Reducing the number of admissions

The mean cost of stay for the inappropriate admissions was US$ 2,532, and for the appropriate US$ 5,800. The observed mean gain in QALY was 2.3. The median bed-occupancy rate in the study period was 0.84 (interquartile range 0.79 to 0.89).

Table 5 shows the effects in terms of costs saved and QALY lost from excluding admissions predicted to be inappropriate. For elective admissions, 9% of the total costs would have been saved and 5% of the total QALY lost. For electively admitted men, 10% (95% CI 5 to 17) of the admissions would have been excluded, 5% (95% CI 2 to 9) of the costs saved and 2% (95% CI 0 to 9) of the QALY lost (n=102). For electively admitted women, the corresponding percentages were 34 (95% CI 22 to 48), 17 (95% CI 6 to 39) and 12 (95% CI 2 to 33)(n=50).

The cost savings from denying care to inappropriate emergency admissions would have been 14% and QALY losses 18% (Table 5).

Sensitivity analysis

The effect of better predictions of appropriateness was explored. To obtain a best-case scenario, the most beneficial admissions among those which had been falsely classified as inappropriate were reclassified as appropriate, and the most costly among those which had been falsely classified as appropriate were reclassified as inappropriate. Assuming that the sensitivity could only be improved slightly from the observed 88 to 90%, but that the specificity could increase from 28 to 50%, 6 and 23 patients, respectively, would need to be reclassified. Under these assumptions, US$ 11,983 was saved per QALY lost (Table 6).

Savings and health losses were also estimated under the assumptions that the inappropriate admissions had the same cost as the appropriate, and that all gains in QALY had been overestimated by 100%. Finally, when combining these two assumptions with improved sensitivity and specificity, US$ 26,131 was saved per QALY lost (Table 6).

Discussion

At present, any strategy for reducing the number of inappropriate admissions to hospitals would have to involve clinical judgement in one way or another. To explore whether this approach can be used for reducing costs without resulting in unacceptable health losses, we used panels of experienced board-certified specialists to provide a higher level of expertise than the average admitting physician. To ensure that the panels had all relevant data available, a board-certified specialist of internal medicine (B.O.E.) prepared the summaries which were the basis of their assessments. Even so, the panels' predictions of appropriateness were poor. While they were able to correctly identify 88% of the appropriate admissions, only 27% of the inappropriate were detected. If the admissions predicted to be inappropriate had been excluded, significant savings would have been obtained (12%), but at the cost of an almost equal percentage of the total benefit in QALY (14%)(Table 5).

Some difficulty for one panel in predicting the other panel's assessment after discharge would be expected due to inter-observer variation. However, fair agreement between the panels was found in the agreement study of group 1. Uncertainty about diagnosis and effect of treatment at admission was probably the most important explanation for the poor predictions. Presumably, there was insufficient information for making any accurate estimate of the effect of the hospital stays for many of the patients (Table 4). It is difficult to see how this situation could have been improved for emergency admissions, but more information could perhaps have been obtained for elective patients before admission to allow better predictions.

Since one third of these admissions were inappropriate as judged after discharge, the potential for better selection of patients was considerable (Table 5).

Rationing based on the panels' predictions of appropriateness would have saved US$ 3,910 per QALY lost for elective and 1,693 for emergency admissions (Table 5). Since the cost analysis only included costs incurred during the included stays, the savings may have been under-estimated. The reason is that many patients with chronic diseases would subsequently have been treated in other parts of the hospital, e.g. in outpatient clinics and other clinical departments, where more costs would have incurred, partly as a consequence of decisions about follow-up made during the included stays. If the patient had not been admitted in the first place, these costs would have been saved in addition to the costs incurred during the included stay. In some studies, US$ 50,000 per QALY has been used as an upper limit for interventions considered to be cost-effective [25,26], which in the present investigation would correspond to the minimum amount that would have had to be saved per QALY lost. However, even allowing for a substantial under-estimation of costs and over-estimation of gains in QALY, the savings per QALY in the present study would have been lower. In the sensitivity analysis, US$ 26,131 per QALY was the maximum saving attained when assuming both higher sensitivity and specificity, more costly inappropriate admissions and lower gains in QALY than observed(Table 6).

One possibility for improving the panels' predictions could have been to give a more detailed specification of the alternatives to hospital care. This approach was chosen by Coast et al who considered 12 alternatives to admission to a department of

general medicine and geriatrics. However, although an alternative was found for 20%, few resources were saved by exploiting this potential [5].

An interesting finding of this study was that rationing of elective admissions would have had different effects for the two sexes. Few resources would have been saved and few QALYs lost for men, whereas a 17% cost reduction would have been obtained at the cost of a 12% loss in QALY for women. The logistic regression analysis indicated that this effect was independent of diagnosis. The difference was not caused by different predictive abilities for the two sexes, as this would have been shown by a significant interaction term between gender and observed appropriateness. This result suggests that reducing admissions in this manner might have discriminated women.

Most previous studies of inappropriate admissions have relied on the Appropriateness Evaluation Protocol (AEP) and similar instruments [27-29]. The AEP has also been used in a major effort to assess the extent of inappropriate health care in the European Union [30]. The main differences between the AEP and our method were that i) the AEP partly relies on information that is only available after admission, thus precluding its use for *predicting* inappropriate admissions, and ii) that it is a screening tool which has been validated against expert clinical judgement, whereas we used clinical judgement directly for evaluating the admissions [27,31,32]. Our results question the assumption that this instrument could reduce hospital costs to any significant degree without leading to health losses. The same applies to other forms of utilisation review, which in the USA have been shown to reduce both the number

of admissions and costs [10,11,33-35]. None of these studies includes an assessment of how this affects the quality and outcome of care.

Two limitations of the cost analysis should be noted. First, we were not able to calculate marginal costs, *i.e.* the cost of treating one more patient, which are most relevant for estimating potential savings. Since the department operated below full capacity, the savings obtained would have been lower than our estimates. Second, the cost analysis was made from the hospital's perspective. A societal perspective would have been preferable, but the task of estimating societal costs for the heterogeneous group of patients in a department of internal medicine would have been insurmountable. It can be assumed that many patients would have been treated elsewhere if not admitted, and that the societal savings would have been lower than our estimates. Accordingly, a cost analysis without these limitations would probably have supported our findings.

We conclude that, in the investigated department, reducing the number of admissions based on predictions of appropriateness at admission would have resulted in unacceptable health losses relative to the savings. The extent to which this conclusion can be generalised is uncertain, but it indicates that a high rate of inappropriate admissions does not necessarily imply that cost savings of the same magnitude can be obtained. In the absence of evidence to the contrary, decision makers should not implement programs to reduce inappropriate admissions without considering their effects on both costs and health benefits.

References

1. Eddy DM. Clinical decision making: from theory to practice. Rationing resources while improving quality. How to get more for less. **JAMA** 1994; **272**: 817-824.

2. Brook RH. Appropriateness: the next frontier. **BMJ** 1994; **308**: 218-219.

3. Berwick DM. Eleven worthy aims for clinical leadership of health system reform. **JAMA** 1994; **272**: 797-802.

4. Payne SM. Identifying and managing inappropriate hospital utilization: a policy synthesis. **Health Serv Res** 1987; **22**: 709-769.

5. Coast J, Inglis A, Morgan K, Gray S, Kammerling M, Frankel S. The hospital admissions study in England: are there alternatives to emergency hospital admission? **J Epidemiol Community Health** 1995; **49**: 194-199.

6. Bentes M, Gonsalves ML, Santos M, Pina E. Design and development of a utilization review program in Portugal. **Int J Qual Health Care** 1995; **7**: 201-212.

7. Lorenzo S, Sunol R. An overview of Spanish studies on appropriateness of hospital use. **Int J Qual Health Care** 1995; **7**: 213-218.

8. Fellin G, Apolone G, Tampieri A, et al. Appropriateness of hospital use: an overview of Italian studies. **Int J Qual Health Care** 1995; **7**: 219-225.

9. Lang T, Davido A, Logerot H, Meyer L. Appropriateness of admissions: the French experience. **Int J Qual Health Care** 1995; **7**: 233-238.

10. Wickizer TM, Wheeler JR, Feldstein PJ. Does utilization review reduce unnecessary hospital care and contain costs? **Med Care** 1989; **27**: 632-647.

11. Scheffler RM, Sullivan SD, Ko TH. The impact of Blue Cross and Blue Shield Plan utilization management programs, 1980-1988. **Inquiry** 1991; **28**: 263-275.

12. Shapiro MF, Wenger NS. Rethinking utilization review. **N Engl J Med** 1995; **333**: 1353-1354.

13. Eriksen BO, Kristiansen IS, Nord E, et al. The cost of inappropriate admissions: A study of health benefits and resource utilization in a department of internal medicine. **J Intern Med** 1999; (In Press)

14. Eriksen BO, Almdahl SM, Hensrud A, et al. Assessing health benefit from hospitalization: Agreement between expert panels. **Int J Technol Assess Health Care** 1996; **12**: 126-135.

15. Mehrez A, Gafni A. Quality-adjusted life years, utility theory, and healthy-years equivalents. **Med Decis Making** 1989; **9**: 142-149.

16. Johannesson M. QALYs, HYEs and individual preferences--a graphical illustration. **Soc Sci Med** 1994; **39**: 1623-1632.

17. Torrance GW, Feeny D. Utilities and quality-adjusted life years. **Int J Technol Assess Health Care** 1989; **5**: 559-575.

18. Johannesson M. The ranking properties of healthy-years equivalents and quality-adjusted life-years under certainty and uncertainty. **Int J Technol Assess Health Care** 1995; **11**: 40-48.

19. Eriksen BO, Kristiansen IS, Nord E, et al. Does admission to a department of internal medicine improve patient quality of life?. **J Intern Med** 1998; **244**: 397-404.

20. Eriksen BO, Kristiansen IS, Nord E, et al. Does admission to a medical department improve patient life expectancy?. **J Clin Epidemiol** 1997; **50**: 987-995.

21. Drummond MF, O'Brien B, Stoddart GL, Torrance GW. Methods for the Economic Evaluation of Health Care Programmes. Oxford: Oxford Medical Publications; 1997.

22. Efron B, Tibshirani RJ. An Introduction to the Bootstrap. New York: Chapman & Hall; 1993.

23. Cohen J. Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. **Psychol Bull** 1968; **70**: 213-220.

24. Fleiss JL. Statistical Methods of Rates and Proportions. New York: John Wiley & Sons Inc; 1981.

25. Hamel MB, Phillips RS, Davis RB, et al. Outcomes and cost-effectiveness of initiating dialysis and continuing aggressive care in seriously ill hospitalized adults. SUPPORT Investigators. Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments. **Ann Intern Med** 1997; **127**: 195-202.

26. Kuntz KM, Tsevat J, Goldman L, Weinstein MC. Cost-effectiveness of routine coronary angiography after acute myocardial infarction. **Circulation** 1996; **94**: 957-965.

27. Strumwasser I, Paranjpe NV, Ronis DL, Share D, Sell LJ. Reliability and validity of utilization review criteria. Appropriateness Evaluation Protocol, Standardized Medreview Instrument, and Intensity-Severity-Discharge criteria. **Med Care** 1990; **28**: 95-111.

28. Mozes B, Rosenblum Y, Rom L, Friedman N, Shabtai E, Porat A. Medical patients assessment protocol: a tool for evaluating the appropriateness of utilizing hospital-stay days for acute medical patients; development, reliability and applications. **Am J Med Qual** 1996; **11**: 18-24.

29. Restuccia JD. The evolution of hospital utilization review methods in the United States. **Int J Qual Health Care** 1995; **7**: 253-260.

30. Liberati A, Apolone G, Lang T, Lorenzo S. A European project assessing the appropriateness of hospital utilization: background, objectives and preliminary results. The Project Steering Group and the Coordinating Center. **Int J Qual Health Care** 1995; **7**: 187-199.

31. Gertman PM, Restuccia JD. The appropriateness evaluation protocol: a technique for assessing unnecessary days of hospital care. **Med Care** 1981; **19**: 855-871.

32. Siu AL, Sonnenberg FA, Manning WG, et al. Inappropriate use of hospitals in a randomized trial of health insurance plans. **N Engl J Med** 1986; **315**: 1259-1266.

33. Restuccia JD. The effect of concurrent feedback in reducing inappropriate hospital utilization. **Med Care** 1982; **20**: 46-62.

34. Feldstein PJ, Wickizer TM, Wheeler JR. Private cost containment. The effects of utilization review programs on health care use and expenditures. **N Engl J Med** 1988; **318**: 1310-1314.

35. Schwartz WB, Mendelson DN. Hospital cost containment in the 1980s. Hard lessons learned and prospects for the 1990s. **N Engl J Med** 1991; **324**: 1037-1042.

Acknowledgements

This study was supported by grants from the Norwegian Medical Association Funds for Quality Improvement and the University Hospital of Tromsø Research Fund. We are grateful to R. Robertsen and G. Thorsen for participating in the expert panels.

Legends

Figure 1 Inclusion of patients and assessment of appropriateness in the Tromsø Medical Department Health Benefit Study

Table 1 Predictions and observations of appropriateness of consecutive admissions to a department of internal medicine

	Observations after discharge		Total (%)
	Inappropriate admissions (%)	Appropriate admissions (%)	
Predictions at admission			
Inappropriate admissions	28 (27[1])	38 (12)	66 (16)
Appropriate admissions	74 (73)	282 (88[2])	356 (84)
Total	102 (100)	320 (100)	422 (100)

[1] Specificity, [2] Sensitivity

Table 2 Sensitivity and specificity of the expert panels' prediction that an admissions would be judged appropriate after discharge

	ICD9-code	N	Sensitivity (%) (95% CI)[1]		Specificity (%) (95% CI)[1]	
Total		422	88	(85 to 92)	27	(19 to 37)
Sex:						
Men		262	93	(89 to 96)	24	(13 to 35)
Women		160	81	(74 to 88)	33	(19 to 49)
Age group:						
<50 years		93	89	(81 to 96)	32	(15 to 50)
50-69 years		180	90	(85 to 95)	17	(5 to 30)
>=70 years		149	85	(78 to 91)	35	(21 to 51)
Admission category:						
Elective		152	89	(83 to 95)	31	(19 to 43)
Emergency		270	88	(84 to 92)	24	(13 to 36)
Diagnostic category:						
Infectious disease	001-139	17	85	(60 to 100)	25	(0 to 100)
Malignant disease	140-208	42	88	(76 to 97)	10	(0 to 33)
Endocrinological disease	240-259	11	75	(43 to 100)	67	(0 to 100)
Acute myocardial infarction	410	30	96	(86 to 100)	0	-
Angina pectoris	411-414	85	97	(93 to 100)	29	(7 to 53)
Other heart disease	420-429	45	90	(80 to 98)	25	(0 to 100)
Cerebrovascular disease	430-438	21	94	(81 to 100)	50	(0 to 100)
Pneumonia and influenza	480-487	16	83	(57 to 100)	25	(0 to 100)
Chronic obstr. pulm. disease	496	20	88	(70 to 100)	50	(0 to 100)
Heapatobiliary/pancreatic disease	570-579	13	100	-	33	(0 to 80)
Undiagnosed symptoms	780-789	30	71	(43 to 93)	19	(0 to 40)
Other		92	80	(70 to 89)	35	(15 to 55)

[1] bias-corrected and accelerated 95% confidence intervals calculated with the bootstrap algorithm from 1,000 runs of a Monte Carlo simulation

Table 3 Multivariate logistic regression analysis of the prediction that an elective admission will be judged appropriate after discharge (n=152)

Independent variables	Parameter estimate	P	Odds ratio	95% confidence interval
Intercept	1.36	0.24		
Sex (0=male, 1=female)	-1.36	0.01	0.26	(0.09 to 0.74)
Age in years divided by 10	-0.01	0.94	0.99	(0.71 to 1.37)
Disease category				
Infectious diseases	-1.27	0.39	0.28	(0.02 to 5.17)
Malignant diseases	0.02	0.97	1.02	(0.24 to 4.33)
Endocrinological diseases	-1.18	0.26	0.31	(0.04 to 2.38)
Angina pectoris	0.57	0.44	1.77	(0.42 to 7.49)
Other heart diseases	0.25	0.79	1.29	(0.21 to 8.04)
Heapatobiliary/pancreatic diseases	0.38	0.70	1.46	(0.21 to 9.95)
Undiagnosed symptoms	-0.08	0.93	0.92	(0.14 to 6.17)
Other	0.00	Reference	1.00	
Observed appropriateness after discharge (0=inappropriate, 1=appropriate)	1.31	0.01	3.72	(1.41 to 9.79)

Chi-square for covariates 25.15, d.f. 10, p=0.005

Table 4 The 10 patients with the greatest discrepancy between predicted and observed health benefits

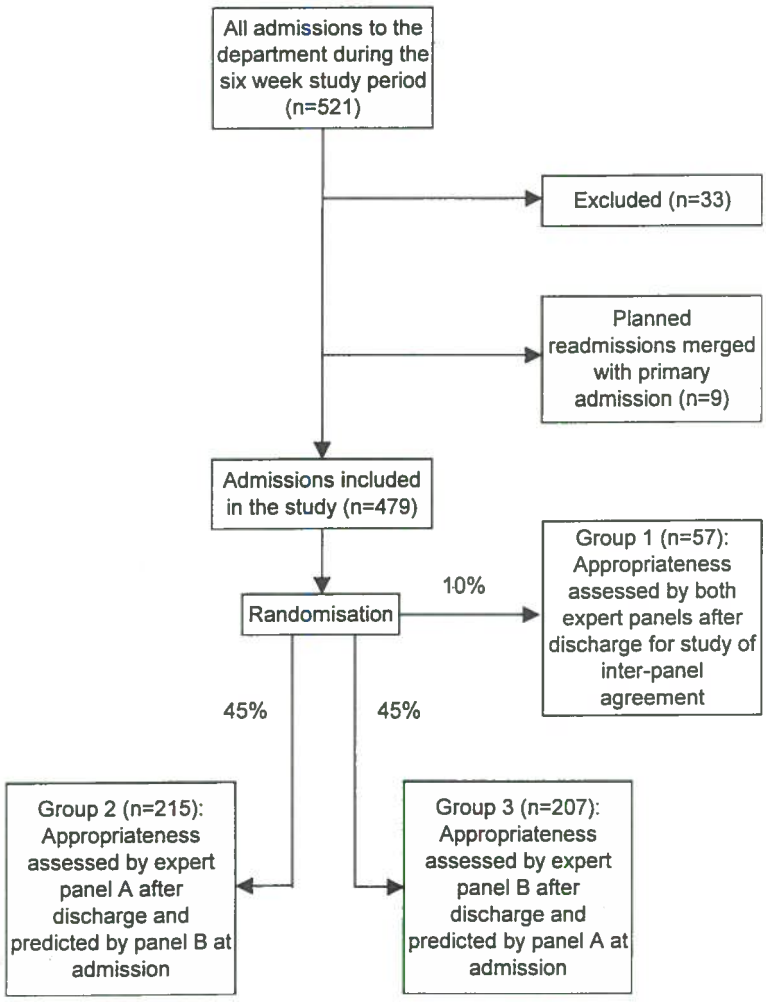
Sex	Age in years	ICD9 code	Appropriateness/ health gain		Total cost in US$	Clinical details
			Predicted	Observed		
Female	28	034	50 QALY1	Inappropriate	1,714	Admitted for suspected meningitis, but was diagnosed to have acute streptococcal tonsillitis. Treated with antibiotics. Judged to could have had the same health benefit without admission.
Male	43	482	29 QALY	Inappropriate	632	Discharged after stay for pneumonia treated with anibiotics 4 days earlier. Readmitted because of fever and slight hemoptysis. An elevated mycoplasma titer was found and he was given doxycycline.
Female	40	427	11 QALY	Inappropriate	611	Paroxystic supraventricular tachycardia. Admitted for magnesium loading as part of workup. Borderline result, magnesium not instituted.
Male	64	785	9 QALY	Inappropriate	17,229	Admitted for septicaemia and multi-organ failure. Treated with mechanical ventilation and hemodialysis. Died after 4 days in hospital
Male	48	412	7 QALY	Inappropriate	1,199	Acute myocardial infarction 3 years previously. Uncharacteristic chest pain. Normal findings on coronary angiography.
Female	28	038	Inappropriate	51 QALY	1,273	Successfully treated septicaemia with group A streptococci. Not diagnosed as septicaemia at admission.
Male	60	250	Inappropriate	13 QALY	1,379	Non-insulin-dependent diabetes mellitus for four years. Now persistent hyperglycemia. HbA1C 18.6, S-glucose 24. Insulin treatment initiated.
Female	42	413	Inappropriate	13 QALY	3,318	Previously considered for possible angina pectoris. Now admitted for chest pain. Coronary angiography showed stenosis of left anterior descending coronary artery and hypokinesia of left ventricle. Aortacoronary bypass planned.
Male	63	038	Inappropriate	12 QALY	8,367	Successfully treated urosepsis
Female	64	427	Inappropriate	8 QALY	3,954	Admitted for atrial fibrillation caused by idiopathic dilated cardiomyopathia. Treated with digitoxin, verapamil and warfarin.

Table 5 Effects of reducing the number of admissions to a department of internal medicine on the basis of predictions of appropriateness made at admission compared to the potential effects if the predictions had been perfect (n=422)

Admission category	Effects of not admitting patients predicted to have been inappropriately admitted				Potential effects if predictions had been perfect	
	Per cent of patients not admitted (95% CI)	Per cent of costs saved (95% CI)	Per cent of gain in QALY lost (95% CI)	Costs saved (US$)/ QALY lost (95% CI)	Per cent of patients not admitted (95% CI)	Per cent of costs saved (95% CI)
Elective (n=152)	18 (12 to 24)	9 (5 to 15)	5 (1 to 12)	3,910 (1,887 to 21,548)	34 (27 to 43)	17 (11 to 26)
Emergency (n=270)	14 (11 to 19)	14 (5 to 26)	18 (6 to 34)	1,693 (474 to 6,525)	19 (14 to 23)	10 (6 to 15)
All (n=422)	16 (12 to 19)	12 (6 to 22)	14 (5 to 27)	1,953 (699 to 5,688)	24 (20 to 28)	12 (9 to 16)

Table 6 Sensitivity analysis of health losses and cost savings from excluding admissions predicted to be inappropriate

	QALY lost	Costs (US$) saved	Costs saved (US$)/ QALY lost
Result of study	134	262,985	1,961
Assumptions in sensitivity analysis			
Sensitivity 90% and specificity 50%	32	381,399	11,983
Cost of appropriate and inappropriate admissions equal	134	330,692	2,466
Over-estimation of all gains in QALY by 100%	67	262,985	3,922
All of the three scenarios above combined	16	415,870	26,131


All admissions to the department during the six week study period (n=521)
Excluded (n=33)
Planned readmissions merged with primary admission (n=9)
Admissions included in the study (n=479)
Randomisation
10%
Group 1 (n=57): Appropriateness assessed by both expert panels after discharge for study of inter-panel agreement
45%
45%
Group 2 (n=215): Appropriateness assessed by expert panel A after discharge and predicted by panel B at admission
Group 3 (n=207): Appropriateness assessed by expert panel B after discharge and predicted by panel A at admission

ISM SKRIFTSERIE - FØR UTGITT:

1. Bidrag til belysning av medisinske og sosiale forhold i Finnmark fylke, med særlig vekt på forholdene blant finskættede i Sør-Varanger kommune.
   Av Anders Forsdahl, 1976. (nytt opplag 1990)

2. Sunnhetstilstanden, hygieniske og sosiale forhold i Sør-Varanger kommune 1869-1975 belyst ved medisinal-beretningene.
   Av Anders Forsdahl, 1977.

3. Hjerte-karundersøkelsen i Finnmark - et eksempel på en populasjonsundersøkelse rettet mot cardiovasculære sykdommer. Beskrivelse og analyse av etterundersøkelsesgruppen.
   Av Jan-Ivar Kvamme og Trond Haider, 1979.

4. The Tromsø Heart Study: Population studies of coronary risk factors with special emphasis on high density lipoprotein and the family occurrence of myocardial infarction.
   Av Olav Helge Førde og Dag Steinar Thelle, 1979.

5. Reformer i distriktshelsetjenesten III: Hypertensjon i distriktshelsetjenesten.
   Av Jan-Ivar Kvamme, 1980.

6. Til professor Knut Westlund på hans 60-års dag, 1983.

7.* Blodtrykksovervåkning og blodtrykksmåling.
   Av Jan-Ivar Kvamme, Bernt Nesje og Anders Forsdahl, 1983.

8.* Merkesteiner i norsk medisin reist av allmennpraktikere - og enkelte utdrag av medisinalberetninger av kulturhistorisk verdi.
   Av Anders Forsdahl, 1984.

9. "Balsfjordsystemet." EDB-basert journal, arkiv og statistikksystem for primærhelsetjenesten.
   Av Toralf Hasvold, 1984.

10. Tvunget psykisk helsevern i Norge. Rettsikkerheten ved slikt helsevern med særlig vurdering av kontrollkommisjonsordningen.
    Av Georg Høyer, 1986.

11. The use of self-administered questionnaires about food habits. Relationships with risk factors for coronary heart disease and associations between coffee drinking and mortality and cancer incidence.
    Av Bjarne Koster Jacobsen, 1988.

12.* Helse og ulikhet. Vi trenger et handlingsprogram for Finnmark.
    Av Anders Forsdahl, Atle Svendal, Aslak Syse og Dag Thelle, 1989.

13. Health education and self-care in dentistry - surveys and interventions.
Av Anne Johanne Søgaard, 1989.

14. Helsekontroller i praksis. Erfaringer fra prosjektet helsekontroller i Troms 1983-1985.
Av Harald Siem og Arild Johansen, 1989.

15. Til Anders Forsdahls 60-års dag, 1990.

16. Diagnosis of cancer in general practice. A study of delay problems and warning signals of cancer, with implications for public cancer information and for cancer diagnostic strategies in general practice.
Av Knut Holtedahl, 1991.

17. The Tromsø Survey. The family intervention study. Feasibility of using a family approach to intervention on coronary heart disease. The effect of lifestyle intervention of coronary risk factors.
Av Synnøve Fønnebø Knutsen, 1991.

18. Helhetsforståelse og kommunikasjon. Filosofi for klinikere.
Av Åge Wifstad, 1991.

19. Factors affecting self-evaluated general health status - and the use of professional health care services.
Av Knut Fylkesnes, 1991.

20. Serum gamma-glutamyltransferase: Population determinants and diagnostic characteristics in relation to intervention on risk drinkers.
Av Odd Nilssen, 1992.

21. The Healthy Faith. Pregnancy outcome, risk of disease, cancer morbidity and mortality in Norwegian Seventh-Day-Adventists.
Av Vinjar Fønnebø, 1992.

22. Aspects of breast and cervical cancer screening.
Av Inger Torhild Gram, 1992.

23. Population studies on dyspepsia and peptic ulcer disease: Occurrence, aetiology, and diagnosis. From The Tromsø Heart Study and The Sørreisa Gastrointestinal Disorder Studie.
Av Roar Johnsen, 1992.

24. Diagnosis of pneumonia in adults in general practice.
Av Hasse Melbye, 1992.

25. Relationship between hemodynamics and blood lipids in population surveys, and effects of n-3 fatty acids.
Av Kaare Bønaa, 1992.

26. Risk factors for, and 13-year mortality from cardiovascular disease by socioeconomic status. A study of 44690 men and 17540 women, ages 40-49. **Av Hanne Thürmer, 1993.**

27. Utdrag av medisinalberetninger fra Sulitjelma 1891-1990. **Av Anders Forsdahl, 1993.**

28. Helse, livsstil og levekår i Finnmark. Resultater fra Hjerte-karundersøkelsen i 1987-88. Finnmark III. **Av Knut Westlund og Anne Johanne Søgaard, 1993.**

29. Patterns and predictors of drug use. A pharmacoepidemiologic study, linking the analgesic drug prescriptions to a population health survey in Tromsø, Norway. **Av Anne Elise Eggen, 1994.**

30. ECG in health and disease. ECG findings in relation to CHD risk factors, constitutional variables and 16-year mortality in 2990 asymptomatic Oslo men aged 40-49 years in 1972. **Av Per G. Lund-Larsen, 1994.**

31. Arrhythmia, electrocardiographic signs, and physical activity in relation to coronary heart risk factors and disease. The Tromsø Study. **Av Maja-Lisa Løchen, 1995.**

32. The Military service: mental distress and changes in health behaviours among Norwegian army conscript. **Av Edvin Schei, 1995.**

33. The Harstad injury prevention study: Hospital-based injury recording and community-based intervention. **Av Børge Ytterstad, 1995.**

34.* Vilkår for begrepsdannelse og praksis i psykiatri. En filosofisk undersøkelse. **Av Åge Wifstad, 1996.** (utgitt Tano Aschehoug forlag 1997)

35. Dialog og refleksjon. Festskrift til professor Tom Andersen på hans 60-års dag, 1996.

36. Factors affecting doctors´ decision making. **Av Ivar Sønbø Kristiansen, 1996.**

37. The Sørreisa gastrointestinal disorder study. Dyspepsia, peptic ulcer and endoscopic findings in a population. **Av Bjørn Bernersen, 1996.**

38. Headache and neck or shoulder pain. An analysis of musculoskeletal problems in three comprehensive population studies in Northern Norway. **Av Toralf Hasvold, 1996.**

39. Senfølger av kjernefysiske prøvespreninger på øygruppen Novaya Semlya i perioden 1955 til 1962. Rapport etter programmet "Liv". Arkangelsk 1994.
Av A.V. Tkatchev, L.K. Dobrodeeva, A.I. Isaev, T.S. Podjakova, 1996.

40. Helse og livskvalitet på 78 grader nord. Rapport fra en befolkningsstudie på Svalbard høsten 1988.
Av Helge Schirmer, Georg Høyer, Odd Nilssen, Tormod Brenn og Siri Steine, 1997.

41. Physical activity and risk of cancer. A population based cohort study including prostate, testicular, colorectal, lung and breast cancer.
Av Inger Thune, 1997.

42. The Norwegian - Russian Health Study 1994/95. A cross-sectional study of pollution and health in the border area.
Av Tone Smith-Sivertsen, Valeri Tchachtchine, Eiliv Lund, Tor Norseth, Vladimir Bykov, 1997.

43. Use of alternative medicine by Norwegian cancer patients
Av Terje Risberg, 1998.

44. Incidence of and risk factors for myocardial infarction, stroke, and diabetes mellitus in allmenn general population. The Finnmark Study 1974-1989.
Av Inger Njølstad, 1998.

45. General practitioner hospitals: Use and usefulness. A study from Finnmark County in North Norway.
Av Ivar Aaraas, 1998.

45B Sykestuer i Finnmark. En studie av bruk og nytteverdi.
Av Ivar Aaraas, 1998.

46. No går det på helsa laus. Helse, sykdom og risiko for sykdom i to nord-norske kystsamfunn.
Av Jorid Andersen, 1998.

47. The Tromsø Study: Risk factors for non-vertebral fractures in a middle-aged population.
Av Ragnar Martin Joakimsen, 1999.

De som er merket med * har vi dessverre ikke flere eksemplar av.